

Fiducial, confidence and objective Bayesian posterior distributions for a multidimensional parameter

Piero Veronese and Eugenio Melilli
Bocconi University, Milano, Italy

Abstract

We propose a way to construct fiducial distributions for a multidimensional parameter using a step-by-step conditional procedure related to the inferential importance of the components of the parameter. For discrete models, in which the non-uniqueness of the fiducial distribution is well known, we propose to use the geometric mean of the “extreme cases” and show its good behavior with respect to the more traditional arithmetic mean. Connections with the generalized fiducial inference approach developed by Hannig and with confidence distributions are also analyzed. The suggested procedure strongly simplifies when the statistical model belongs to a subclass of the natural exponential family, called conditionally reducible, which includes the multinomial and the negative-multinomial models. Furthermore, because fiducial inference and objective Bayesian analysis are both attempts to derive distributions for an unknown parameter without any prior information, it is natural to discuss their relationships. In particular, the reference posteriors, which also depend on the importance ordering of the parameters are the natural terms of comparison. We show that fiducial and reference posterior distributions coincide in the location-scale models, and we characterize the conditionally reducible natural exponential families for which this happens. The discussion of some classical examples closes the paper.

Keywords: Confidence distribution, Jeffreys prior, Location-scale parameter model, Multinomial model, Natural exponential family, Reference prior.

1 Introduction

Fiducial distributions, after having been introduced by Fisher (1930, 1935) and widely discussed (and criticized) in the subsequent years, have been de facto brushed aside for a long time and only recently they have obtained new vitality. The original idea of Fisher was to construct a *distribution* for a parameter which includes all the information given by the data, without resorting to the Bayes theorem. This is obtained by transferring the randomness from the observed quantity given by the statistical model to the parameter. Originally Fisher considered a continuous sufficient statistic S with distribution function F_θ , depending on a real parameter θ . Let $q_\alpha(\theta)$ denote the quantile of order α of F_θ and

let s be a realization of S . If $q_\alpha(\theta)$ is increasing in θ (i.e., F_θ is decreasing in θ), the statement $s < q_\alpha(\theta)$ is equivalent to $\theta > q_\alpha^{-1}(s)$ and thus Fisher assumes $q_\alpha^{-1}(s)$ as the quantile of order $1 - \alpha$ of a distribution which he names *fiducial*. The set of all quantiles $q_\alpha^{-1}(s)$, $\alpha \in (0, 1)$, establishes the fiducial distribution function $H_s(\theta)$ so that

$$H_s(\theta) = 1 - F_\theta(s) \quad \text{and} \quad h_s(\theta) = \frac{\partial}{\partial \theta} H_s(\theta) = -\frac{\partial}{\partial \theta} F_\theta(s). \quad (1)$$

Of course H_s , and its density h_s , must be properly modified if F_θ is increasing in θ .

Fisher (1973, cap.VI) also provides some examples of multivariate fiducial distributions obtained by a “step-by-step” procedure, but he never develops a general and rigorous theory. This fact, along with the problem to cover discrete models, the presence of some inconsistencies of the fiducial distribution (e.g. the marginalization paradox, see Dawid & Stone, 1982), and the difficulties in its interpretation, gave rise to a quite strong negative attitude towards Fisher proposal.

In the renewed interest for the fiducial approach a relevant role is played by the *generalized fiducial inference* introduced and developed by Hannig (2009, 2013), see also Hannig et al. (2016) for a review. He provides a formal and mathematically rigorous definition which has a quite general applicability. The crucial element of his definition is a data-generating equation $\mathbf{X} = \mathbf{G}(\mathbf{U}, \boldsymbol{\theta})$, which links the unknown parameter $\boldsymbol{\theta}$ and the observed data \mathbf{X} through a random element \mathbf{U} having a known distribution. Roughly speaking, by shifting the randomness of \mathbf{U} from \mathbf{X} to $\boldsymbol{\theta}$ (inverting \mathbf{G} with respect to $\boldsymbol{\theta}$ after having fixed $\mathbf{X} = \mathbf{x}$), the distribution given by the statistical model leads to a distribution for the parameter $\boldsymbol{\theta}$. Contrary to the original idea of Fisher, the generalized fiducial distribution is non-unique and Hannig widely discusses this point. Applications to different statistical models can be found for instance in Hannig et al. (2007), Hannig & Iyer (2008) and Wandler & Hannig (2012).

Other recent contributions to the topic of fiducial distributions are given by Taraldsen & Lindqvist (2013), Martin & Liu (2013) and Veronese & Melilli (2015), henceforth V&M (2015). In this last paper the authors derive fiducial distributions for a parameter in a discrete or continuous real natural exponential family (NEF), and discuss some of their properties with particular emphasis on the frequentist coverage of the fiducial intervals.

In the past fiducial distributions have often been associated with *confidence distributions* even if these latter have a different meaning. A modern definition of confidence distribution is given in Schweder & Hjort (2002) and Singh et al. (2005), see the book by Schweder & Hjort (2016) for a complete and updated review on confidence distributions and their connections with fiducial inference. It is important to emphasize that a confidence distribution must be regarded as a function of the data with reasonable properties from a purely frequentist point of view. A confidence distribution is conceptually similar

to a point estimator: as there exist several unbiased estimators, several confidence distributions can be provided for the same parameter and choosing among them can be done resorting to further optimality criteria. Thus the confidence distribution theory allows to compare, in a quite general setting, formal distributions for the parameter derived by different statistical procedures.

In this paper we suggest a way to construct a unique distribution for a multidimensional parameter, indexing discrete and continuous models, following a step-by-step procedure similar to that used by Fisher (1973) in some examples. We call it *fiducial distribution*, but we look at it simply as a distribution on the parameter space in the spirit of the confidence distribution theory. The key-point of the construction is the procedure by conditioning: the distribution of the data is factorized as a product of one-dimensional laws and, for each of these, the fiducial density for a real parameter component, possibly conditional on other components, is obtained. The joint fiducial density for the parameter is then defined as the product of the (conditional) one-dimensional fiducial densities. It is well known that Fisher’s fiducial argument presents several drawbacks in higher dimensions, essentially because one cannot recover the fiducial distribution for a function of the parameters starting from the joint fiducial distribution, see Schweder & Hjort (2016, Ch. 6 and 9). Our approach, when it can be applied, presents the advantage to construct sequentially the fiducial distribution directly on the parameters of interest and different fiducial distributions can be obtained focusing on different parameters of interest. Also, it should be noticed that a general definition of confidence distribution for a multidimensional parameter does not exist and more attention is given to the construction of approximate confidence curves for specific nested families of regions, see Schweder & Hjort (2016, Ch. 9 and Sec. 15.4).

Interestingly, our joint fiducial distribution coincides in many cases with the Bayesian posterior obtained using the reference prior. This fact motivates the second goal of the paper: to investigate the relationships between the objective Bayesian posteriors and the suggested fiducial distributions. Objective Bayesian analysis, see e.g. Berger (2006), essentially studies how to perform a *good* Bayesian inference, especially for moderate sample size, when one is unwilling or unable to assess a subjective prior. Under this approach, the prior distribution is derived directly from the model and thus it is labeled as *objective*. The reference prior, introduced by Bernardo (1979) and developed by Berger & Bernardo (1992), is the most successful default prior proposed in the literature. For a multidimensional parameter the reference prior depends on the grouping and ordering of its components and, in general, no longer coincides with the Jeffreys prior. This is the reference prior only for a real parameter and it is unsatisfactory otherwise, as well known.

Lindley (1958) was the first to discuss the connections between fiducial and posterior distributions for a real parameter, when a real continuous sufficient statistic exists. V&M (2015) extend this result to real discrete NEFs, characterizing all families admitting a *fiducial prior*, i.e. a prior leading to a posterior coinciding with the fiducial distribution. This prior is strictly related to the Jeffreys prior. We show here that when the parameter is multidimensional this relationship no longer holds and a new one is established with the reference prior. In particular we prove results for location-scale parameter models and *conditionally reducible* NEFs, a subclass of NEFs defined in Consonni & Veronese (2001).

The paper is structured as follows. Section 2 reviews some basic facts on fiducial and confidence distributions for real NEFs and on generalized fiducial distributions. The proposal for constructing a step-by-step multivariate fiducial distribution is presented in Section 3, which also discusses: the relationships with confidence distributions (Section 3.1), the use of the geometric mean of fiducial densities for solving the non-uniqueness problem in discrete models (Section 3.2), the connections with the generalized fiducial inference and the consistency with the sufficiency principle (Section 3.3). Section 3.4 studies the fiducial distributions for conditionally reducible NEFs and provides their explicit expression for a particular subclass which includes the multinomial and the negative-multinomial model. Section 4 analyzes the relationships between the fiducial distributions and the reference posteriors, in particular for location-scale parameter models (Section 4.1) and NEFs (Section 4.2), characterizing those which admit the fiducial prior. Section 5 discusses further examples in which fiducial and reference posteriors coincide. Section 6 concludes the paper presenting some possible asymptotic extensions. Finally, Appendix A1 collects some useful technical results on conditionally reducible NEFs, while Appendix A2 includes the proofs of all the results stated in the paper.

2 Preliminary results

The modern definition of *confidence distribution* for a real parameter ϕ of interest, see Schweder & Hjort (2002, 2016) and Singh et al. (2005), can be formulated as follows:

Definition 1. Let $\{F_{\phi, \boldsymbol{\lambda}}, \phi \in \Phi \subseteq \mathbb{R}, \boldsymbol{\lambda} \in \Lambda\}$ be a parametric model for data $\mathbf{X} \in \mathcal{X}$; here ϕ is the parameter of interest and $\boldsymbol{\lambda}$ is a nuisance parameter. A function $C : \mathcal{X} \times \Phi \rightarrow \mathbb{R}$ is a confidence distribution for ϕ if $C(\mathbf{x}, \cdot)$ is a distribution function for each $\mathbf{x} \in \mathcal{X}$ and $C(\mathbf{X}, \phi_0)$ has a uniform distribution in $(0, 1)$ under $F_{\phi_0, \boldsymbol{\lambda}_0}$, where $(\phi_0, \boldsymbol{\lambda}_0)$ is the true parameter value.

The relevant requirement in the previous definition is the uniformity of the distribution, which ensures the correct coverage of the confidence intervals. As seen in Section

1, the confidence distribution theory must be placed in a purely frequentist context and allows to compare distributions on the parameter space, obtained using different approaches. Finally, the definition of confidence distribution can be generalized by requiring that the uniformity assumption holds only asymptotically.

Strictly linked to the notion of confidence distribution is that of *confidence curve*, defined, for each observed $\mathbf{X} = \mathbf{x}$, as the function $\phi \rightarrow cc(\phi) = |1 - 2C(\mathbf{x}, \phi)|$; see Schweder & Hjort (2016). This function gives the extremes of equal-tail confidence intervals for any level $1 - \alpha$, allowing a fast and clear comparison of confidence distributions with respect to their interval length. When the parameter of interest is multidimensional, how to extend the definitions of confidence distribution and confidence curve is much less clear and various proposals have been made, see Schweder & Hjort (2002, 2016) and Singh et al. (2005).

As detailed in Section 1, Hannig (2009) has proposed the notion of *generalized fiducial distribution*, which is based on a data-generating equation $\mathbf{X} = \mathbf{G}(\boldsymbol{\theta}, \mathbf{U})$. Because several functions \mathbf{G} can generate the same statistical model, and not all the resulting fiducial distributions are reasonable in terms of properties or computational tractability, Hannig (2013, Sec. 5) gives some hints on the choice of a *default* function \mathbf{G} . In particular, if $\mathbf{X} = (X_1, \dots, X_n)$ is an independent identically distributed (i.i.d.) random sample from an (absolutely) continuous distribution function $F_{\boldsymbol{\theta}}$, with density $f_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \mathbb{R}^d$, he suggests to use $X_i = F_{\boldsymbol{\theta}}^{-1}(U_i)$, $i = 1, \dots, n$, where U_i are i.i.d. uniform random variables on $(0, 1)$ and $F_{\boldsymbol{\theta}}^{-1}$ is the inverse (or generalized inverse) of $F_{\boldsymbol{\theta}}$. If other regularity assumptions are satisfied, the generalized fiducial distribution for $\boldsymbol{\theta}$ can be written as

$$r(\boldsymbol{\theta}) = \frac{f_{\boldsymbol{\theta}}(\mathbf{x})J(\mathbf{x}, \boldsymbol{\theta})}{\int_{\Theta} f_{\boldsymbol{\theta}}(\mathbf{x})J(\mathbf{x}, \boldsymbol{\theta})d\boldsymbol{\theta}}, \quad (2)$$

where the expression of $J(\mathbf{x}, \boldsymbol{\theta})$, given in Hannig (2013, formula (3.7)), is

$$J(\mathbf{x}, \boldsymbol{\theta}) = \sum_{\{(i_1, \dots, i_d): 1 \leq i_1 < \dots < i_d \leq n\}} \left| \frac{\det \left(\frac{d}{d\boldsymbol{\theta}} (F_{\boldsymbol{\theta}}(x_{i_1}), \dots, F_{\boldsymbol{\theta}}(x_{i_d})) \right)}{\prod_{j=1}^d f_{\boldsymbol{\theta}}(x_{i_j})} \right|. \quad (3)$$

In (3) the numerator of the ratio is the determinant of the matrix whose kj -entry is $\partial F_{\boldsymbol{\theta}}(x_{i_j}) / \partial \theta_k$. This procedure leads to the Fisher definition of fiducial density (1) when $n = d = 1$.

Hannig (2013, Example 4) explicitly recognizes the advise of Wilkinson (1977) that the choice of a fiducial distribution should depend on the parameter of interest and uses the well known example of d independent normal distributions $N(\mu_i, 1)$, in which the parameter of interest is $\boldsymbol{\theta} = (\sum_{i=1}^d \mu_i^2)^{1/2}$. He shows that the default data-generating equations $X_i = \mu_i + U_i$, $i = 1, \dots, d$, lead to a fiducial distribution which has good frequentist properties for inference on the μ 's, but very bad ones when the interest is on

θ , as already recognized by Stein (1959). Thus Hannig suggests an ad hoc alternative equation, which leads to a better solution. Notice that our general procedure, suggested in the next section, constructs a fiducial distribution starting directly from the parameter of interest and do not required the choice a priori of a data generating function.

Fiducial distributions and their properties, with particular emphasis on the frequentist coverage of the fiducial intervals, for a discrete or a continuous real regular NEF, are discussed in V&M (2015). More specifically, consider the sufficient statistic S associated with a sample of size n and denote by \mathcal{S} its support. Let $F_\theta(s)$ be the distribution function of S and $p_\theta(s) = \exp\{\theta s - nM(\theta)\}$ the corresponding density (with respect to a measure ν). Let $a = \inf \mathcal{S}$, $b = \sup \mathcal{S}$ and define $\mathcal{S}^* = [a, b]$ if $\nu(a) > 0$, otherwise $\mathcal{S}^* = (a, b)$. Then, for $s \in \mathcal{S}^*$, Petrone & Veronese (2010) have proved that

$$H_s(\theta) = \begin{cases} 0 & \theta \leq \inf \Theta \\ 1 - F_\theta(s) & \inf \Theta < \theta < \sup \Theta \\ 1 & \theta \geq \sup \Theta \end{cases} \quad (4)$$

is a fiducial distribution function for the natural parameter θ . It follows that the fiducial density of θ is

$$h_s(\theta) = \frac{\partial}{\partial \theta} H_s(\theta) = -\frac{\partial}{\partial \theta} F_\theta(s) = \int_{(-\infty, s]} (nM'(\theta) - t)p_\theta(t) d\nu(t). \quad (5)$$

It is important to underline, and simple to verify, that the distribution function H_s is also a confidence distribution (only asymptotically, in the discrete case), according to Definition 1.

Notice that, for discrete NEFs, $F_\theta(s) = \Pr_\theta\{S \leq s\}$ and $\Pr_\theta\{S < s\}$ do not coincide and thus, besides H_s in (4), one could define a *left fiducial distribution* as

$$H_s^\ell(\theta) = 1 - \Pr_\theta\{S < s\}. \quad (6)$$

For convenience, sometimes H_s will be called *right fiducial distribution*. A standard way to overcome this non-uniqueness is referring to the half-correction device (see Schweder & Hjort, 2016, pag. 62) which amounts to consider the mixture $H_s^A(\theta) = (H_s(\theta) + H_s^\ell(\theta))/2 = \Pr_\theta\{S > s\} + \Pr_\theta\{S = s\}/2$, whose density is the arithmetic mean of $h_s(\theta)$ and $h_s^\ell(\theta)$. Instead, we will suggest to average h_s and h_s^ℓ using their *geometric mean* h_s^G (suitably normalized) and show that it presents better properties than h_s^A (Section 3.2) and a more direct connection with objective Bayesian inference (Section 4.2), even if, operationally, the difference is usually not particularly big.

Table 1 provides the fiducial distributions obtained in V&M (2015) for some important discrete and continuous NEFs, which will be used in the forthcoming examples. It also establishes the abbreviations used in the paper for the standard distributions.

Table 1: *Fiducial distributions for some real NEFs*

	Sufficient statistic	Fiducial distributions
$N(\mu, \sigma^2)$ (σ^2 known)	$S = \sum_i X_i$	$H_s(\mu) : N(s/n, \sigma^2/n)$
$N(\mu, \sigma^2)$ (μ known)	$S = \sum_i (X_i - \mu)^2$	$H_s(\sigma^2) : \text{In-Ga}(n/2, s/2)$
$\text{Ga}(\alpha, \lambda)$ (α known)	$S = \sum_i X_i$	$H_s(\lambda) : \text{Ga}(n\alpha, s)$
$\text{Pa}(\lambda, x_0)$ (x_0 known)	$S = \sum_i \log(X_i/x_0)$	$H_s(\lambda) : \text{Ga}(n, s)$
$\text{We}(\lambda, c)$ (c known)	$S = \sum_i X_i^c$	$H_s(\lambda) : \text{Ga}(n, s)$
$\text{Bi}(m, p)$ (m known)	$S = \sum_i X_i$	$H_s(p) : \text{Be}(s+1, nm-s)$ $H_s^\ell(p) : \text{Be}(s, nm-s+1)$ $H_s^G(p) : \text{Be}(s+1/2, nm-s+1/2)$
$\text{Po}(\mu)$	$S = \sum_i X_i$	$H_s(\mu) : \text{Ga}(s+1, n)$ $H_s^\ell(\mu) : \text{Ga}(s, n)$ $H_s^G(\mu) : \text{Ga}(s+1/2, n)$
$\text{Ne-Bi}(m, p)$ (m known)	$S = \sum_i X_i$	$H_s(p) : \text{Be}(nm, s+1)$ $H_s^\ell(p) : \text{Be}(nm, s)$ $H_s^G(p) : \text{Be}(nm, s+1/2)$

The following notations are used: $\text{Ga}(\alpha, \lambda)$ for a gamma distribution with shape α and mean α/λ ; $\text{In-Ga}(\alpha, \lambda)$ for an inverse-gamma distribution (if $X \sim \text{Ga}(\alpha, \lambda)$ then $1/X \sim \text{In-Ga}(\alpha, \lambda)$); $\text{Be}(\alpha, \beta)$ for a beta distribution with parameters α and β ; $\text{Bi}(m, p)$ for a binomial distribution with m trials and success probability p ; $\text{Ne-Bi}(m, p)$ for a negative-binomial with m successes and success probability p ; $\text{Po}(\mu)$ for the Poisson distribution with mean μ ; $\text{Pa}(\lambda, x_0)$ for a Pareto distribution with density $\lambda x_0^\lambda x^{-\lambda-1}$, $x > x_0 > 0$, $\lambda > 0$; $\text{We}(\lambda, c)$ for a Weibull distribution with density $c\lambda x^{c-1} \exp(-\lambda x^c)$, $x, \lambda, c > 0$.

3 Fiducial distributions for multidimensional parameters

A *natural* way to construct a suitable fiducial distribution for a multidimensional parameter is to follow the step-by-step procedure used by Fisher (1973) in some examples. The key-point of our proposal stems on the factorization of the sampling distribution as a product of one-dimensional conditional laws. For each of these the fiducial density for a real component of the parameter, possibly conditional on other components, is defined. It is well known that different factorizations of sampling distributions can produce different joint fiducial distributions, see e.g. Dempster (1963). However, we do not consider this aspect a drawback of the procedure if it is linked to the inferential importance ordering of the parameter components implied by the factorization. For example, if a parameter $\boldsymbol{\theta} \in \mathbb{R}^2$ is transformed in such a way that ϕ is the parameter of interest and λ the nuisance, the obvious ordering is (ϕ, λ) and a suitable factorization must be defined accordingly, see Example 4 (ctd.) in this section for an illustration. The crucial role played by the ordering of the parameters accordingly to their inferential importance is widely acknowledged by objective Bayesian inference, in which reference priors are different for different orderings, see Section 4.

In order to construct a fiducial distribution, we consider two basic transformations: one involving the sample data $\mathbf{X} = (X_1, \dots, X_n)$, having a distribution parameterized by $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$, $d \leq n$, and one involving $\boldsymbol{\theta}$. Given \mathbf{X} , consider a statistic $\mathbf{T} = (T_1, \dots, T_m)$, $d \leq m \leq n$, with density $p_{\boldsymbol{\theta}}(\mathbf{t})$, which summarizes \mathbf{X} without losing information on $\boldsymbol{\theta}$. \mathbf{T} can be a sufficient statistic or a one-to-one transformation of \mathbf{X} . Split \mathbf{T} in $(\mathbf{T}_{[d]}, \mathbf{T}_{-[d]})$, where $\mathbf{T}_{[d]} = (T_1, \dots, T_d)$ and $\mathbf{T}_{-[d]} = (T_{d+1}, \dots, T_m)$, and suppose that $\mathbf{T}_{-[d]}$ is ancillary for $\boldsymbol{\theta}$. As a consequence $p_{\boldsymbol{\theta}}(\mathbf{t}) = p_{\boldsymbol{\theta}}(\mathbf{t}_{[d]}|\mathbf{t}_{-[d]})p(\mathbf{t}_{-[d]})$ and all the information on $\boldsymbol{\theta}$ provided by \mathbf{X} are included in the conditional distribution of $\mathbf{T}_{[d]}$ given $\mathbf{T}_{-[d]}$.

Assume now that there exists a one-to-one smooth reparameterization from $\boldsymbol{\theta}$ to $\boldsymbol{\phi}$, with ϕ_1, \dots, ϕ_d ordered with respect to their importance, such that

$$p_{\boldsymbol{\phi}}(\mathbf{t}_{[d]}|\mathbf{t}_{-[d]}) = \prod_{k=1}^d p_{\phi_{d-k+1}}(t_k|\mathbf{t}_{[k-1]}, \mathbf{t}_{-[d]}; \boldsymbol{\phi}_{[d-k]}). \quad (7)$$

The density $p_{\phi_{d-k+1}}(t_k|\mathbf{t}_{[k-1]}, \mathbf{t}_{-[d]}; \boldsymbol{\phi}_{[d-k]})$, with the corresponding distribution function $F_{\phi_{d-k+1}}(t_k|\mathbf{t}_{[k-1]}, \mathbf{t}_{-[d]}; \boldsymbol{\phi}_{[d-k]})$, must be interpreted as the conditional distribution of T_k given $(\mathbf{T}_{[k-1]} = \mathbf{t}_{[k-1]}, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]})$, parameterized by ϕ_{d-k+1} , assuming $\boldsymbol{\phi}_{[d-k]}$ known. In the following, we will always assume that all the one-dimensional conditional distribution functions F_{ϕ_j} 's involved in the analysis are monotone and differentiable in ϕ_j and have limits 0 and 1 when ϕ_j tends to the boundaries of its domain. Notice that this is always true if F_{ϕ_j} belongs to a NEF, see (4). Under these assumptions, the joint fiducial

density of ϕ is obtained as

$$h_{\mathbf{t}}(\phi) = \prod_{k=1}^d h_{\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}}(\phi_{d-k+1} | \phi_{[d-k]}), \quad (8)$$

and

$$h_{\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}}(\phi_{d-k+1} | \phi_{[d-k]}) = \left| \frac{\partial}{\partial \phi_{d-k+1}} F_{\phi_{d-k+1}}(t_k | \mathbf{t}_{[k-1]}, \mathbf{t}_{-[d]}; \phi_{[d-k]}) \right|. \quad (9)$$

Several applications of this procedure to well known models will be provided in Section 5. Here we illustrate some interesting features of the fiducial distribution (8).

i) The existence of an ancillary statistic is not necessary if there exists a sufficient statistic with the same dimension of the parameter ($m = d$). An important case is $m = d = 1$ so that formula (8) and (9) reduce to $h_t(\phi) = |\partial F_\phi(t)/\partial \phi|$, the original formula suggested by Fisher (1930).

ii) If one is only interested in ϕ_1 , it follows from (7) that it is enough to consider

$$h_{\mathbf{t}}(\phi_1) = \left| \frac{\partial}{\partial \phi_1} F_{\phi_1}(t_d | \mathbf{t}_{[d-1]}, \mathbf{t}_{-[d]}) \right|,$$

which, depending on all observations, does not lose any sample information. A typical choice for T_d is given by the maximum likelihood estimator $\hat{\phi}_1$ of ϕ_1 and thus, when $\hat{\phi}_1$ is not sufficient, we have to consider the distribution of $\hat{\phi}_1$ given the ancillary statistic $\mathbf{t}_{-[d]}$. Similarly, if one is interested in ϕ_1, ϕ_2 , it is enough to consider $h_{\mathbf{t}}(\phi_1) \cdot h_{\mathbf{t}_{[d-1]}, \mathbf{t}_{-[d]}}(\phi_2 | \phi_1)$, and so on.

iii) When an ancillary statistic $\mathbf{T}_{-[d]}$ is needed, the fiducial distribution (8) is invariant with respect to any one-to-one transformation of $\mathbf{T}_{-[d]}$. All the sampling distributions are conditional on it and thus any transformation establishes the same constraints; see Section 4.1 for an example.

iv) The construction by successive conditioning makes the fiducial distribution invariant under the so called one-to-one *lower triangular transformation* of $\mathbf{T}_{[d]}$, for fixed $\mathbf{T}_{-[d]}$. More precisely, we consider a transformation $\mathbf{T}^* = (\mathbf{T}_{[d]}^*, \mathbf{T}_{-[d]})$ such that $T_k^* = g_k(\mathbf{T}_{[k]}, \mathbf{T}_{-[d]})$, for $k = 1, \dots, d$. To see this, assuming for instance $t_k^* = g_k(\mathbf{t}_{[k]}, \mathbf{t}_{-[d]})$ increasing in t_k , it is sufficient to show that

$$\begin{aligned} & \Pr_{\phi_{d-k+1}}(T_k^* \leq t_k^* | \mathbf{T}_{[k-1]}^* = \mathbf{t}_{[k-1]}^*, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]}; \phi_{[d-k]}) \\ &= \Pr_{\phi_{d-k+1}}(g_k(\mathbf{T}_{[k]}, \mathbf{T}_{-[d]}) \leq g_k(\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}) | \mathbf{T}_{[k-1]}^* = \mathbf{t}_{[k-1]}^*, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]}; \phi_{[d-k]}) \\ &= \Pr_{\phi_{d-k+1}}(T_k \leq t_k | \mathbf{T}_{[k-1]} = \mathbf{t}_{[k-1]}, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]}; \phi_{[d-k]}). \end{aligned}$$

It follows immediately that \mathbf{T} and \mathbf{T}^* lead to the same fiducial distribution.

v) If $(\mathbf{T}_{[k-1]}, \mathbf{T}_{-[d]})$ is sufficient for $\phi_{[d-k]}$, for each k , then the conditional distribution of T_k given $(\mathbf{T}_{[k-1]} = \mathbf{t}_{[k-1]}, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]})$ does not depend on $\phi_{[d-k]}$ and the fiducial distribution (8) becomes the product of the “marginal” fiducial distributions of the ϕ_k ’s. As a consequence, (9) can be used alone to make inference on ϕ_{d-k+1} and the fiducial distribution does not depend on the inferential ordering of the parameters. An important case in which this happens will be discussed in Section 3.4.

We close this section establishing the invariance property of the fiducial distribution $h_{\mathbf{t}}(\phi)$ under a lower triangular transformation, i.e. a transformation from ϕ to $\lambda = (\lambda_1, \dots, \lambda_d)$, say, which maintains the same decreasing ordering of importance in the components of the two vectors.

Proposition 1. *If $\phi = \phi(\lambda)$ is a one-to-one lower triangular continuously differentiable function from Λ to Φ , then the fiducial distribution $h_{\mathbf{t}}^{\phi}(\phi)$, obtained applying (8) to the model $p_{\phi}(\mathbf{t})$, and the fiducial distribution $h_{\mathbf{t}}^{\lambda}(\lambda)$, obtained applying (8) to the model $p_{\lambda}(\mathbf{t}) = p_{\phi(\lambda)}(\mathbf{t})$, are such that, for each measurable $A \subset \Phi$,*

$$\int_A h_{\mathbf{t}}^{\phi}(\phi) d\phi = \int_{\lambda^{-1}(A)} h_{\mathbf{t}}^{\lambda}(\lambda) d\lambda. \quad (10)$$

3.1 Relationships with confidence distributions

Given a real NEF, $H_s(\theta)$ in (4) is an exact or approximate confidence distribution if the observations are continuous or discrete, respectively. It is possible to verify that the same is true for the marginal fiducial distribution of the main parameter of interest ϕ_1 in the more general definition (8). Indeed, the distribution function of ϕ_1 is $H_{\mathbf{t}}(\phi_1) = 1 - F_{\phi_1}(t_d | \mathbf{t}_{[d-1]}, \mathbf{t}_{-[d]})$, so that the first requirement in Definition 1 is clearly satisfied, thanks to the assumption on the distribution function given after formula (7). For what concerns the uniformity condition, assuming that F_{ϕ_1} is decreasing in ϕ_1 (if it is increasing, replace $1 - F_{\phi_1}$ with F_{ϕ_1}), we have, for $u \in (0, 1)$ and arbitrary ϕ ,

$$\begin{aligned} \Pr_{\phi} \left(H_{\mathbf{t}_{[d]}, \mathbf{t}_{-[d]}}(\phi_1) \leq u \right) &= 1 - \Pr_{\phi} \left(F_{\phi_1}(t_d | \mathbf{T}_{[d-1]}, \mathbf{T}_{-[d]}) < 1 - u \right) = \\ &= 1 - \int \Pr_{\phi} \{ F_{\phi_1}(t_d | \mathbf{t}_{[d-1]}, \mathbf{t}_{-[d]}) < 1 - u \} dF_{\phi}(\mathbf{t}_{[d-1]}, \mathbf{t}_{-[d]}) = u \end{aligned}$$

because, by construction, the integrand is equal to $1 - u$ for all fixed $(\mathbf{t}_{[d-1]}, \mathbf{t}_{-[d]})$.

3.2 The discrete case: the geometric mean of the left and right fiducial densities

As mentioned in Section 2, for a discrete statistic S with distribution depending on a real parameter θ , we suggest to use the geometric mean of the right and left fiducial

densities, $h_s^G(\theta) = c^{-1}(h_s(\theta)h_s^\ell(\theta))^{1/2}$, where c is the normalizing constant, instead of their arithmetic mean $h_s^A(\theta)$.

A first justification of the use of the geometric mean of densities is suggested by Berger et al. (2015) who mention its property to be the density “closest” to h_s and h_s^ℓ with respect to the the Kullback-Leibler divergence, as specified in the following proposition. We give a simple proof of this fact, without resorting to the calculus of variations. Recall that, given two densities p and q , having the same support and the same dominating measure ν , the Kullback-Leibler divergence of p from q is defined as $KL(q|p) = \int q(x) \log(q(x)/p(x)) d\nu(x)$.

Proposition 2. *Consider two densities p_1 and p_2 with the same support. The density q which minimizes $KL(q|p_1) + KL(q|p_2)$ is given by $q = p^G \propto (p_1 p_2)^{1/2}$, which is the (normalized) geometric mean of p_1 and p_2 .*

Furthermore, Krishnamoorthy & Lee (2010) observe that a distribution for θ , whose aim is to give a synthesis of two fiducial distributions, should “stochastically” lie between them. In our setting, the extreme distributions are H_s and H_s^ℓ . This property is surely satisfied by the arithmetic mean, because $H_s(\theta) < H_s^A(\theta) < H_s^\ell(\theta)$ uniformly with respect to θ , for each s belonging to the set \mathcal{S}_0 for which both H_s and H_s^ℓ can be defined. The same inequalities are true for H_s^G under mild assumptions. As usual, here we assume that $H_s(\theta)$ is defined as $1 - F_\theta(s)$.

Proposition 3. *Let p_θ , $\theta \in \Theta \subseteq \mathbb{R}$, be the probability mass function of a real observation S , having a continuous derivative with respect to θ . For each $s \in \mathcal{S}_0$, assume that the function*

$$\gamma_s(\theta) = \left(\frac{\partial p_\theta(s)}{\partial \theta} \right) / \left(-\frac{\partial F_\theta(s)}{\partial \theta} \right) = \left(\frac{\partial p_\theta(s)}{\partial \theta} \right) / h_s(\theta)$$

is decreasing on Θ . Then $H_s(\theta) < H_s^G(\theta) < H_s^\ell(\theta)$ uniformly on Θ .

The assumptions required in the previous proposition are satisfied by many important models. For example we have the following

Corollary 1. *If p_θ is the probability mass function of a real NEF, then $H_s(\theta) < H_s^G(\theta) < H_s^\ell(\theta)$ uniformly on Θ .*

We now discuss the relationship between H_s^G and H_s^A .

Proposition 4. *Let p_θ , $\theta \in \Theta \subseteq \mathbb{R}$, be the probability mass function of a real observation S , satisfying the following assumptions in addition to those stated in Proposition 3:*

$$\lim_{\theta \rightarrow \inf \Theta} \gamma_s(\theta) = +\infty; \quad \lim_{\theta \rightarrow \sup \Theta} \gamma_s(\theta) = -1.$$

Then, for each $s \in \mathcal{S}_0$, there exists $\theta^* \in \Theta$ (depending on s) such that $H_s^G(\theta) < H_s^A(\theta)$ for $\theta < \theta^*$ and $H_s^G(\theta) > H_s^A(\theta)$ for $\theta \geq \theta^*$.

The result in Proposition 4 is important in connection with confidence intervals, because it shows that H_s^G gives, for a fixed level, a confidence interval smaller than that obtained from $H_s^A(\theta)$; see Figure 1 (graph 2) for an example.

Notice that the assumptions on $\gamma_s(\theta)$ in Proposition 4 are fulfilled by a real NEF with natural parameter space $\Theta = \mathbb{R}$, as it occurs in the binomial and Poisson models. However, these assumptions are not necessary to ensure the stated behavior of H_s^G and H_s^A , that we conjecture to be quite general, as the following example shows.

Example 1. Consider an i.i.d. sample of size n from a logarithmic distribution with parameter $\theta \in (0, 1)$ with probability mass function

$$p_\theta(x) = \frac{\theta^x}{-x \log(1 - \theta)} I_{\{1, 2, \dots\}}(x).$$

The sufficient statistic $T = \sum_{i=1}^n X_i$ is distributed as

$$p_\theta(t) = \frac{n! |s(t, n)| \theta^t}{t! (-\log(1 - \theta))^n} I_{\{n, n+1, \dots\}}(t),$$

where $s(t, n)$ is the Stirling number of the first kind with arguments t and n , see Johnson et al. (2005). The distribution of T belongs to a real NEF with $F_\theta(t)$ decreasing in θ , so that the fiducial distribution function H_t , for $t = n, n+1, \dots$ and $\theta \in (0, 1)$, is

$$H_t(\theta) = 1 - F_\theta(t) = 1 - \sum_{j=n}^t \frac{n! |s(j, n)| \theta^j}{j! (-\log(1 - \theta))^n}.$$

For this model

$$\gamma_t(\theta) = - \left(\frac{\partial p_\theta(s)}{\partial \theta} \right) / \left(\frac{\partial F_\theta(s)}{\partial \theta} \right) = - \frac{|s(t, n)| \theta^{t-1} (n\theta + t(1 - \theta) \log(1 - \theta))}{t! \sum_{j=n}^t |s(j, n)| \theta^{j-1} (n\theta + j(1 - \theta) \log(1 - \theta)) / j!}.$$

It can be seen that, for each $t \geq n$, γ_t is decreasing in θ and

$$\lim_{\theta \rightarrow 0^+} \gamma_t(\theta) = +\infty, \quad \lim_{\theta \rightarrow 1^-} \gamma_t(\theta) = - \left(\sum_{j=n}^t \frac{|s(j, n)|}{|s(t, n)|} \frac{t!}{j!} \right)^{-1} \in (-1, 0).$$

Nevertheless, the fiducial distributions H_t^G and H_t^A behave as stated in Proposition 4, see Figure 1 (graph 1).

Finally, we justify our preference for H_s^G versus H_s^A showing that its confidence risk under quadratic penalty, as defined in Schweder & Hjort (2016, Sec. 5.3), is uniformly better for all the important discrete models reported in Table 1. The confidence risk

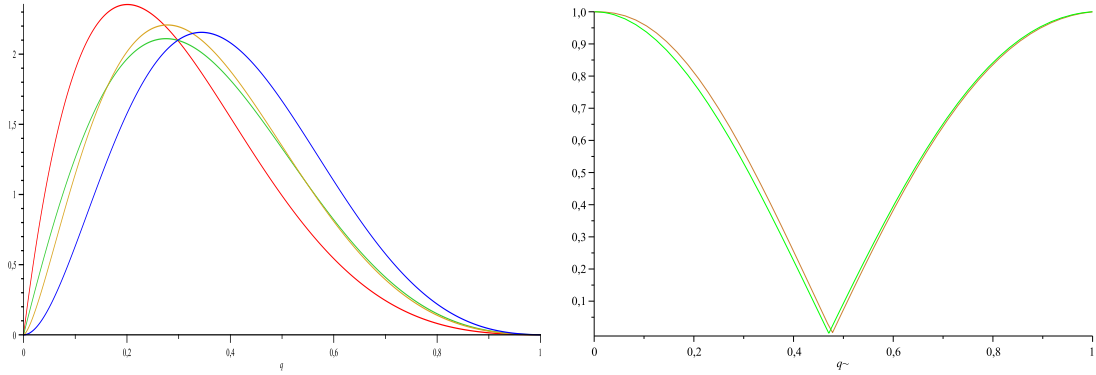


Figure 1: Graph 1: Fiducial distributions for a sample from the logarithmic distribution ($n = 10$, $t = 12$): $h_t^\ell(\theta)$ (red), $h_t^A(\theta)$ (green), $h_t^G(\theta)$ (yellow), $h_t(\theta)$ (blue). Graph 2: Confidence curves for $h_t^A(\theta)$ (green) and $h_t^G(\theta)$ (yellow).

$R(\mu, H_s)$ for the mean parameter μ and a confidence (or fiducial) distribution H_s under quadratic penalty is

$$R(\mu, H_s) = \int (\mu' - \mu)^2 dH_s(\mu') = E_\mu(\text{Var}^{H_s}(\mu)) + E_\mu(\hat{\mu} - \mu)^2,$$

where $\text{Var}^{H_s}(\mu)$ denotes the variance of μ under H_s , E_μ the expected value with respect to the distribution of S and $\hat{\mu} = E^{H_s}(\mu)$ is the mean of μ under H_s . Now, recalling that for the binomial and the negative-binomial distribution in Table 1, assuming $m = 1$ for simplicity, we have $\mu = p$ and $\mu = (1 - p)/p$, it is easy to verify that for both these models and the Poisson model $\hat{\mu}$ is the same under H_s^G and H_s^A . As a consequence, $R(\mu, H_s^A) - R(\mu, H_s^G) = E_\mu(\text{Var}^{H_s^A}(\mu)) - E_\mu(\text{Var}^{H_s^G}(\mu))$ which becomes $(4(n+1)(n+2))^{-1}$, $(4(n-1)(n-2))^{-1}$ and $(4n^2)^{-1}$, for the three models above, respectively. All these values are strictly positive for each n uniformly in μ .

Let us now consider the fiducial distribution for a multivariate parameter defined in (8). For each discrete component of the product, starting from $\Pr_{\phi_{d-k+1}}\{T_k \leq t_k | \mathbf{T}_{[k-1]} = \mathbf{t}_{[k-1]}, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]}; \phi_{[d-k]}\} = F_{\phi_{d-k+1}}(t_k | \mathbf{t}_{[k-1]}, \mathbf{t}_{-[d]}; \phi_{[d-k]})$ and $\Pr_{\phi_{d-k+1}}\{T_k < t_k | \mathbf{T}_{[k-1]} = \mathbf{t}_{[k-1]}, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]}; \phi_{[d-k]}\}$, it is possible to define a right and a left fiducial distribution, respectively, and hence their geometric and arithmetic means. Notice that each component of (8) involves a one-dimensional parameter and a real observation (the remaining quantities being fixed), so that the Propositions 3 and 4 can be applied. Multivariate fiducial distributions for discrete observations can thus be obtained combining in the various possible way these univariate distributions. In particular, we will consider $H_t(\phi)$, obtained as the product of all the *right* univariate conditional fiducial distributions, $H_t^\ell(\phi)$, obtained as the product of all the *left* univariate conditional fiducial distributions, $H_t^A(\phi)$, defined as the product of the d mixtures $H_{\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}}^A = (H_{\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}} + H_{\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}}^\ell)/2$ and finally

$H_{\mathbf{t}}^G(\phi)$, corresponding to the density $h_{\mathbf{t}}^G(\phi)$ obtained as the product of the d geometric means $h_{\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}}^G \propto (h_{\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}} \cdot h_{\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}}^\ell)^{1/2}$. Notice that $h_{\mathbf{t}}^G(\phi)$ coincides with the geometric mean of all the 2^d fiducial densities derived as described above.

3.3 Fiducial inference and the sufficiency principle

The step-by-step procedure introduced at the beginning of Section 3 gives a generalized fiducial distribution, according to Hannig (2009), if one considers as data-generating equation $\mathbf{T} = \mathbf{G}(\phi, \mathbf{U})$ with

$$T_k = \begin{cases} G_k(\phi, \mathbf{U}_{[k]}, \mathbf{U}_{-[d]}) & k = 1, \dots, d \\ U_k & k = d+1, \dots, m \end{cases},$$

where \mathbf{U} is a random vector with a completely known distribution. The functions G_k can be explicitly obtained iteratively as follows:

$$\begin{aligned} T_1 &= G_1(\phi, U_1, \mathbf{U}_{-[d]}) = F_{\phi_d}^{-1}(U_1 | \mathbf{U}_{-[d]}; \phi_{[d-1]}) \\ T_2 &= G_2(\phi, U_1, U_2, \mathbf{U}_{-[d]}) = F_{\phi_{d-1}}^{-1}(U_2 | G_1(\phi, U_1, \mathbf{U}_{-[d]}), \mathbf{U}_{-[d]}; \phi_{[d-2]}) \end{aligned}$$

and so on.

It is interesting to observe that the generalized fiducial distribution $r(\theta)$ given in (2) does not necessarily satisfy the sufficiency principle. This can be verified immediately looking at the Example 2 in Hannig (2013), in which a uniform distribution on (θ, θ^2) is considered and $r(\theta)$ does not depend on the X_i 's only through the sufficient statistic $\mathbf{S} = (X_{(1)}, X_{(n)})$, where $X_{(i)}$ denotes the i -th order statistic. Despite its simple form, this model is highly irregular, but the inconsistency with the sufficiency principle of the generalized fiducial distribution $r(\theta)$ can also occur for more standard models. In particular, if a real continuous sufficient statistic S for a real parameter exists, one could derive two different fiducial distributions starting from S or from the whole sample. A simple example of this issue can be easily constructed considering a beta model with parameters 2 and θ . Another interesting example is the following.

Example 2. Let $\mathbf{X} = (X_1, \dots, X_n)$ be an i.i.d. sample from a truncated exponential density $p_\theta(x) = \theta e^{-\theta x} / (1 - e^{-\theta})$, $0 < x < 1$, $\theta \in \mathbb{R} - \{0\}$. This density is not defined for $\theta = 0$, but it can be completed by continuity setting $p_0(x) = 1$. The distribution function of X_i is $F_\theta(x_i) = (1 - e^{-\theta x_i}) / (1 - e^{-\theta})$, $0 < x_i < 1$, so that from (3) we have

$$J(\mathbf{x}, \theta) = \frac{s}{\theta} + \frac{e^{-\theta}}{\theta(1 - e^{-\theta})} \sum_{i=1}^n (1 - e^{-\theta x_i}),$$

where $s = \sum_{i=1}^n x_i$. Thus, using (2), we obtain

$$r(\theta) \propto \frac{\theta^{n-1}}{(1 - e^{-\theta})^{n+1}} e^{-\theta s} \left(s(1 - e^{-\theta}) + e^{-\theta} \sum_{i=1}^n (1 - e^{-\theta x_i}) \right), \quad \theta \in \mathbb{R}, \quad (11)$$

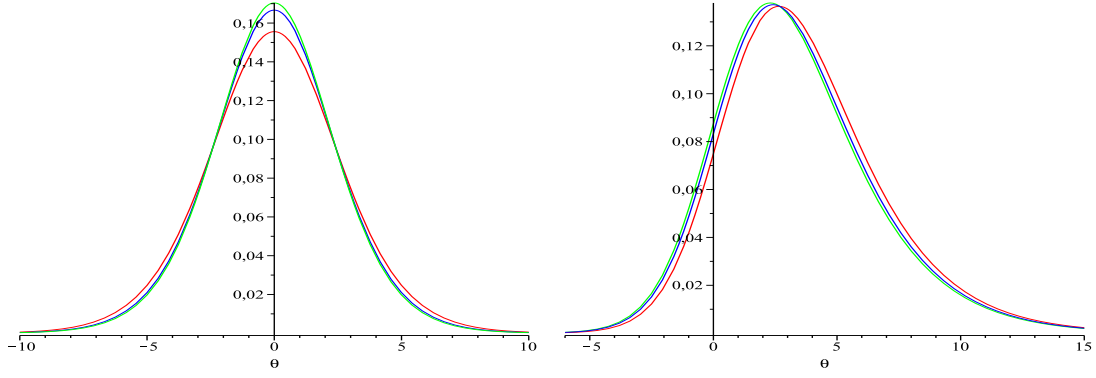


Figure 2: Graph 1: Fiducial densities: $r(\theta; x_1 = 0.05, x_2 = 0.95)$ (red); $r(\theta; x_1 = 0.5, x_2 = 0.5)$ (green); $h(\theta; s = 1)$ (blue). Graph 2: Fiducial densities $r(\theta; x_1 = 0.02, x_2 = 0.48)$ (red); $r(\theta; x_1 = 0.2, x_2 = 0.3)$ (green); $h(\theta; s = 0.5)$ (blue).

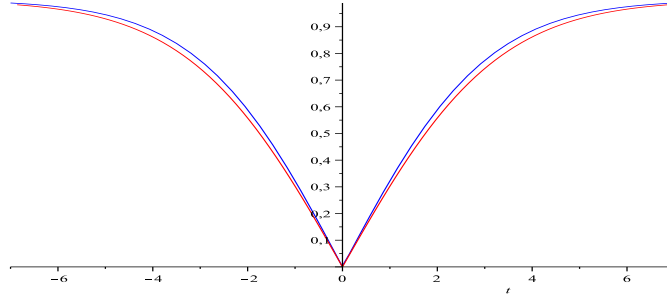


Figure 3: Confidence curves for $r(\theta; x_1 = 0.5, x_2 = 0.5)$ (red) and $h(\theta; s = 1)$ (blue).

which depends on the values of the specific x_i 's. Consider now the sufficient statistic $S = \sum_{i=1}^n X_i$ and, for simplicity, assume $n = 2$. The density of S is

$$p_\theta(s) = \begin{cases} \frac{\theta^2}{(1-e^{-\theta})^2} e^{-ts} s & 0 < s \leq 1 \\ \frac{\theta^2}{(1-e^{-\theta})^2} e^{-ts} (2-s) & 1 < s < 2 \end{cases}.$$

and the generalized fiducial density (11) reduces to $h_s(\theta) = \partial F_\theta(s)/\partial \theta$. In Figure 2 we report the fiducial densities r and h_s for different values of (x_1, x_2) and $s = x_1 + x_2$. For $s = 1$ all densities are symmetric with the mode in 0, while the dispersion is increasing in $|x_1 - x_2|$, so that the more concentrated fiducial density is obtained for $x_1 = x_2 = 0.5$. However, for $s \neq 1$, the densities have different modes and are shifted to the left when x_1 increases. In all cases the fiducial density h_s is in the middle of the various cases. Notice that h_s has all the good properties discussed in V&M (2015) and, in particular, it is a confidence distribution because the model belongs to a NEF. The confidence intervals corresponding to $h_s(\theta)$ are slightly smaller than those corresponding to $r(\theta)$, as can be seen from the confidence curves reported in Figure 3. For instance, when $x_1 = x_2 = 0.5$, the 95% confidence intervals are $(-4.191, 4.191)$ and $(-4.399, 4.399)$, for $h_s(\theta)$ and $r(\theta)$, respectively.

The computation of the fiducial distribution $H_{\mathbf{t}}$ defined in (8) is greatly simplified starting with the sufficient statistic instead of the whole sample. However, when both the alternatives are feasible, they seem to lead to the same result. In particular, the following proposition states that the sufficiency principle is always satisfied by $H_{\mathbf{t}}$ when there exists a complete sufficient statistic for the parameter.

Proposition 5. *Consider the fiducial distribution $H_{\mathbf{t}}(\phi)$, defined in (8) and (9), with \mathbf{T} a one-to-one transformation of the data $\mathbf{X} = (X_1, \dots, X_n)$. If $\mathbf{S} = \mathbf{S}(\mathbf{T})$ is a complete and sufficient statistic of dimension d for ϕ , such that $\mathbf{S} = \mathbf{g}(\mathbf{T}_{[d]}, \mathbf{T}_{-[d]})$ is a one-to-one lower triangular transformation of $\mathbf{T}_{[d]}$ for fixed $\mathbf{T}_{-[d]}$, then the fiducial distribution $H_{\mathbf{s}}(\phi)$ for ϕ , obtained using \mathbf{S} instead of \mathbf{T} in (8) and (9), coincides with $H_{\mathbf{t}}(\phi)$.*

Notice that the completeness of \mathbf{S} is not necessary to satisfy the sufficiency principle, as the following example shows.

Example 3. Given an i.i.d. sample \mathbf{X} of size n from a uniform distribution on $(\theta, \theta + 1)$, it is immediate to verify that the sufficient statistic $\mathbf{S} = (X_{(1)}, X_{(n)})$ is not complete. Because θ is a location parameter, $Z = X_{(n)} - X_{(1)}$ is an ancillary statistic and the fiducial distribution for θ can be obtained starting from the distribution function of $X_{(n)}$ given Z , which is $F_{\theta}(x_{(n)}|z) = (x_{(n)} - z - \theta)/(1 - z)$, $x_{(n)} - 1 < \theta < x_{(n)} - z = x_{(1)}$. Thus

$$h_{\mathbf{s}}(\theta) = -\frac{\partial}{\partial \theta} F_{\theta}(x_{(n)}|z) = -\frac{\partial}{\partial \theta} \frac{x_{(n)} - z - \theta}{1 - z} = \frac{1}{1 - z}, \quad x_{(n)} - 1 < \theta < x_{(n)} - z = x_{(1)}. \quad (12)$$

If we start directly with \mathbf{X} , we can consider the distribution function of X_n given $\mathbf{Z} = (Z_1, \dots, Z_{n-1})$, where $Z_i = X_n - X_i$. Omitting tedious calculations, we have

$$F_{\theta}(x_n | \mathbf{z}) = \frac{x_n - \theta - \max(z_i, 0)}{1 + \min(z_i, 0) - \max(z_i, 0)}, \quad \theta + \max(z_i, 0) < x_n < \theta + 1 + \min(z_i, 0),$$

and thus, for $x_n - 1 - \min(z_i, 0) < \theta < x_n - \max(z_i, 0)$,

$$h_{\mathbf{x}}(\theta) = -\frac{\partial}{\partial \theta} F_{\theta}(x_n | \mathbf{z}) = \frac{1}{1 + \min(z_i, 0) - \max(z_i, 0)}.$$

Observing that $\min(z_i, 0) = z_i$ unless $x_n = x_{(n)}$ and recalling that $z_i = x_n - x_i$, $i = 1, \dots, n-1$, it follows that $x_n - 1 - \min(z_i, 0) = x_{(n)} - 1$ and similarly $x_n - \max(z_i, 0) = x_{(1)}$, so that $h_{\mathbf{x}}(\theta)$ coincides with $h_{\mathbf{s}}(\theta)$ given in (12).

3.4 Conditionally reducible natural exponential families

Consider a multivariate natural exponential family whose density, with respect to a fixed σ -finite positive measure ν , is given by

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \exp \left\{ \sum_{k=1}^d \theta_k x_k - M(\boldsymbol{\theta}) \right\}, \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_d) \in \boldsymbol{\Theta}, \quad \mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d. \quad (13)$$

A NEF is *d-conditionally reducible* (in the sequel cr-NEF) if its joint density (13) can be factorized as a product of d conditional densities each belonging to a real exponential family. More precisely if

$$p_{\phi}(\mathbf{x}|\phi(\boldsymbol{\theta})) = \prod_{k=1}^d p_{\phi_k}(x_k|\mathbf{x}_{[k-1]}; \phi_k(\boldsymbol{\theta})) = \prod_{k=1}^d \exp \{ \phi_k(\boldsymbol{\theta})x_k - M_k(\phi_k(\boldsymbol{\theta}); \mathbf{x}_{[k-1]}) \}, \quad (14)$$

where $\phi = (\phi_1, \dots, \phi_d)$ is a one-to-one function from $\boldsymbol{\Theta}$ onto $\phi(\boldsymbol{\Theta}) = \boldsymbol{\Phi}$. Furthermore, it can be shown that $\boldsymbol{\Phi} = \Phi_1 \times \dots \times \Phi_d$, with $\phi_k \in \Phi_k, k = 1, \dots, d$, so that the ϕ_k 's are variation independent. Notice that ϕ_k is the natural parameter of the k -th conditional distribution. For details on these families, with emphasis on *enriched* conjugate priors and on reference Bayesian analysis, see Consonni & Veronese (2001) and Consonni et al. (2004), respectively. Both these papers deal in particular with the families having *simple quadratic variance function*, named NEF-SQVFs, which include, as most interesting cases, the multinomial and negative-multinomial models, see Casalis (1996) and Appendix A1.

Example 4 (Multinomial model). Consider a random vector \mathbf{X} distributed according to a multinomial distribution and denote by p_k the probability of the k -th outcome $X_k, k = 1, \dots, d$, with $\sum_{k=1}^d x_k \leq N$, and $\sum_{k=1}^d p_k \leq 1$. It is well known that the conditional distribution of X_k given $\mathbf{X}_{[k-1]} = \mathbf{x}_{[k-1]}, k = 2, \dots, d$, is $\text{Bi}(N - \sum_{j=1}^{k-1} x_j, p_k / (1 - \sum_{j=1}^{k-1} p_j))$, whereas the marginal distribution of X_1 is $\text{Bi}(N, p_1)$. Since a binomial distribution is a real NEF, one can factorize the multinomial distribution as in (14) with

$$\phi_k = \log \frac{p_k}{1 - \sum_{j=1}^k p_j}, \quad M_k(\phi_k; \mathbf{x}_{[k-1]}) = \left(N - \sum_{j=1}^{k-1} x_j \right) \log(1 + e^{\phi_k}), \quad \phi_k \in \mathbb{R}. \quad (15)$$

For models belonging to a cr-NEF, the construction of the fiducial distribution proposed in Section 3 drastically simplifies. The existence of a sufficient statistic of the same dimension of the parameter makes the ancillary statistic not necessary, while the ϕ -parameterization, indexing each conditional distribution with a real parameter, implies the independence of the ϕ_k 's under the fiducial distribution.

Proposition 6. *Let \mathbf{S} be the sufficient statistic distributed according to a regular cr-NEF on \mathbb{R}^d , parameterized by $\phi \in \boldsymbol{\Phi} = \Phi_1 \times \dots \times \Phi_d$, with Φ_k coinciding with the natural parameter space of the k -th conditional distribution. Then, for $\mathbf{S} = \mathbf{s}$, with $s_k, k = 1, \dots, d$, satisfying conditions similar to those given before (4),*

$$H_{\mathbf{s}}(\phi) = \prod_{k=1}^d H_{\mathbf{s}_{[k]}}(\phi_k) = \prod_{k=1}^d (1 - F_{\phi_k}(s_k|\mathbf{s}_{[k-1]})) \quad (16)$$

is a fiducial distribution function on ϕ with density

$$h_{\mathbf{s}}(\phi) = \prod_{k=1}^d h_{\mathbf{s}_{[k]}}(\phi_k), \quad \text{where } h_{\mathbf{s}_{[k]}}(\phi_k) = \frac{\partial}{\partial \phi} H_{\mathbf{s}_{[k]}}(\phi_k) = -\frac{\partial}{\partial \phi_k} F_{\phi_k}(s_k | \mathbf{s}_{[k-1]}). \quad (17)$$

The ϕ_k 's are independent under $H_{\mathbf{s}}(\phi)$ and thus their importance ordering is irrelevant. This fact also justifies the simplification in the index notation adopted in (16). Notice, however, that the definition and the interpretation of the ϕ_k 's depend on the particular ordering considered for the X_k 's, as seen in Example 4.

As recalled in Section 2, a general definition of multi-dimensional confidence distribution does not exist. However, in our context, since $H_{\mathbf{s}}(\phi)$ is constructed as a product of *marginal* confidence distributions, it can be considered as a multivariate (possibly asymptotic) confidence distribution for ϕ .

Some of the examples of Section 5 can be reconnected with this framework, but here we consider in specific the NEF-SQVFs, whose variance function is given in (35). For this class, with the exclusion of the negative-multinomial/hyperbolic secant distribution, it is possible to give a simple explicit expression of the fiducial density of ϕ , recalling the definition of $B_k(\phi_k)$ given in (31) and setting $z_k = z_{kk}$ in (35). The specifications of z_{kk} and q , appearing in (35), and of $B_k(\phi_k)$ can be found in Appendix A1.

Proposition 7. *Consider a sample of size n from a Poisson/normal, a multinomial, or a negative-multinomial family on \mathbb{R}^d . If \mathbf{S} denotes the sufficient statistic, then the (right) fiducial distribution for ϕ has density*

$$h_{\mathbf{s}}(\phi) = \prod_{k=1}^d h_{\mathbf{s}_{[k]}}(\phi_k) \propto \prod_{k=1}^d \exp \left\{ \phi_k (s_k + z_k) - \left[n + q \left(\sum_{j=1}^{k-1} s_j - 1 \right) \right] B_k(\phi_k) \right\}, \quad (18)$$

while for the negative-multinomial/gamma/normal family, with an m dimensional negative multinomial component, the (right) fiducial distribution is given by

$$\begin{aligned} h_{\mathbf{s}}(\phi) &= \prod_{k=1}^d h_{[s_k]}(\phi_k) \propto \prod_{k=1}^m \exp\{\phi_k(s_k + 1)\} (1 - \exp(\phi_k))^{n/q + \sum_{j=1}^{k-1} s_j - 1} \\ &\times \exp\{\phi_{m+1} s_{m+1}\} (-\phi_{m+1})^{n/q + \sum_{j=1}^m s_j - 1} \\ &\times \prod_{k=m+2}^d \exp\{\phi_k s_k - n s_{m+1} \phi_k^2 / 2\}. \end{aligned} \quad (19)$$

Notice that the discrete components of a basic NEF-SQVF are integer-valued with $z_k = 1$, so that the left fiducial distribution is obtained by the previous formulas replacing the term $(s_k + 1)$ by s_k in (18) and (19). Thus it follows that the geometric mean $h_{\mathbf{s}}^G$ has the same structure in (18) and (19) with $(s_k + 1/2)$ instead of $(s_k + 1)$.

Example 4 (ctd.). For the multinomial family, because $q = -1/N$, $z_k = 1$ and $B_k(\phi_k) = N \log(1 + e^{\phi_k})$, $k = 1, \dots, d$, it easily follows from formula (18) that

$$h_{\mathbf{s}}^G(\phi) = \prod_{k=1}^d h_{\mathbf{s}_{[k]}}^G(\phi_k) = \prod_{k=1}^d \left\{ \frac{1}{B(s_k + \frac{1}{2}, nN - \sum_{j=1}^k s_j + \frac{1}{2})} \frac{e^{\phi_k(s_k + \frac{1}{2})}}{(1 + e^{\phi_k})^{nN - \sum_{j=1}^{k-1} s_j + 1}} \right\}, \quad (20)$$

where $B(\cdot, \cdot)$ denotes the beta function.

The fiducial distribution for ϕ not always is of particular interest in itself, but it can be used as a starting point for the construction of the fiducial distribution for alternative and more relevant parameters. We consider here the mean-parameter $\boldsymbol{\mu}$, which is a lower triangular transformation of ϕ , see (32), so that its fiducial distribution can be directly obtained from that of ϕ thanks to Proposition 1.

Corollary 2. *The (right) fiducial distribution for the mean parameter $\boldsymbol{\mu}$, relative to the ordering μ_1, \dots, μ_d , for the following NEF-SQVFs on \mathbb{R}^d , has density:*

- *Poisson/normal family (with m Poisson components)*

$$h_{\mathbf{s}}(\boldsymbol{\mu}) \propto \prod_{k=1}^m \mu_k^{s_k-1} \exp(-n\mu_k) \prod_{k=m+1}^d \exp\left\{-\frac{n}{2\sigma^2}(\mu^2 - 2\mu_k s_k/n)\right\}, \quad (21)$$

which corresponds to the product of m densities $Ga(s_k, n)$, $k = 1, \dots, m$ and $(d-m)$ densities $N(s_k/n, \sigma^2/n)$, $k = m+1, \dots, d$.

- *Multinomial family*

$$h_{\mathbf{s}}(\boldsymbol{\mu}) \propto \prod_{k=1}^d \mu_k^{s_k} \left(N - \sum_{j=1}^k \mu_j \right)^{\gamma_k}, \quad (22)$$

where $\gamma_k = -1$ for $k = 1, \dots, d-1$ and $\gamma_d = Nn - 1 - \sum_{j=1}^d s_j$.

- *Negative-multinomial family, with R occurrences in the $(d+1)$ -th cell*

$$h_{\mathbf{s}}(\boldsymbol{\mu}) \propto \prod_{k=1}^d \mu_k^{s_k} \left(R + \sum_{j=1}^k \mu_j \right)^{\gamma_k}, \quad (23)$$

where $\gamma_k = -1$ for $k = 1, \dots, d-1$ and $\gamma_d = -Rn - 1 - \sum_{j=1}^d s_j$.

- *Negative-multinomial/gamma/normal family (with an m -dimensional negative -*

multinomial component with R occurrences in the $(m+1)$ -th cell)

$$\begin{aligned}
h_{\mathbf{s}}(\boldsymbol{\mu}) &\propto \prod_{k=1}^m \mu_k^{s_k} \left(R + \sum_{j=1}^k \mu_j \right)^{\gamma_k} \\
&\times \left(R + \sum_{j=1}^m \mu_j \right)^{Rn + \sum_{j=1}^m s_j} \mu_{m+1}^{-(Rn + \sum_{j=1}^m s_j) - 1} \exp \left\{ -\frac{s_{m+1} \left(R + \sum_{j=1}^m \mu_j \right)}{\mu_{m+1}} \right\} \\
&\times \prod_{k=m+2}^d \mu_{m+1}^{-(d-m-1)} \exp \left\{ -\frac{ns_{m+1}}{2\mu_{m+1}^2} \left(\mu_k^2 - 2\mu_k \mu_{m+1} \frac{s_k}{n} \right) \right\},
\end{aligned} \tag{24}$$

where $\gamma_k = -1$ for $k = 1, \dots, m-1$ and $\gamma_m = -Rn - 1 - \sum_{j=1}^m s_j$. Notice that the density of μ_{m+1} given $\boldsymbol{\mu}_{[m]}$ is an In-Ga($Rn + \sum_{j=1}^m s_j, s_{m+1}(R + \sum_{j=1}^m \mu_j)$), while the density of μ_k given $\boldsymbol{\mu}_{[k-1]}$, $k = m+2, \dots, d$, is a $N(\mu_{m+1}s_k/n, \mu_{m+1}^2/(ns_{m+1}))$ depending only on μ_{m+1} .

Example 4 (ctd.). Inference for the multinomial distribution is usually performed for the cell-probabilities parameter $\mathbf{p} = (p_1, \dots, p_d)$. Since $p_k = \mu_k/N$, the fiducial distribution h^G for \mathbf{p} is easily derived from (22), noting that the left fiducial density can be obtained replacing $(s_k + 1)$ by s_k in (18), (and not in (22), which is derived aggregating the hyperparameters). It follows that the geometric mean $h_{\mathbf{s}}^G(\mathbf{p})$ is given by

$$h_{\mathbf{s}}^G(\mathbf{p}) \propto \prod_{k=1}^d p_k^{s_k - 1/2} \left(1 - \sum_{j=1}^k p_j \right)^{\gamma_k}, \quad \sum_{k=1}^d p_k = 1, \quad 0 < p_k < 1, \tag{25}$$

with $\gamma_k = -1/2$ for $k = 1, \dots, d-1$ and $\gamma_d = Nn - 1/2 - \sum_{j=1}^d s_j$. This is a generalized Dirichlet distribution. Clearly, $h_{\mathbf{s}}^G(\mathbf{p})$ in (25) refers to the specific order of importance p_1, p_2, \dots, p_d . If we change this order, the fiducial distribution will change accordingly.

Similarly, for the negative-multinomial model, with R occurrences in the $(d+1)$ -th cell, $h_{\mathbf{s}}^G(\boldsymbol{\phi})$ can be easily computed from (18) observing that $z_k = 1$, $q = 1/R$ and $B_k(\phi_k) = -R \log(1 - \exp(\phi_k))$.

4 Connections with objective Bayesian inference

As mentioned in Section 1, if we look at fiducial inference as a way to obtain a distribution on the parameter space of the model without any prior information, it appears natural to compare it with objective Bayesian inference. Recall that when a fiducial distribution coincides with a posterior, the corresponding prior is called *fiducial prior*.

The step-by-step construction of the fiducial distribution $h_{\mathbf{t}}(\boldsymbol{\phi})$ defined in (8) is based on the inferential importance ordering of the parameter components ϕ_1, \dots, ϕ_d .

This aspect is also crucial in the procedure adopted to construct reference priors, see Bernardo & Smith (1994, Sec. 5.4.5). The reference prior π^R for a parameter ϕ is generated by successive conditioning, established by the importance ordering of its components, as $\pi^R(\phi) = \prod_{k=1}^d \pi^R(\phi_{d-k+1} | \phi_{[d-k]})$. It is widely recognized that the dependence of the reference prior on the choice of the parameter of interest is necessary to obtain good frequentist properties such as coverage and consistency. For a one-dimensional parameter ϕ the reference prior coincides with the Jeffreys prior $\pi^J(\phi) \propto I(\phi)^{1/2}$, where $I(\phi)$ denotes the Fisher information. While the Jeffreys prior is invariant under a reparameterization of the model, the reference prior (and thus the reference posterior) is generally not invariant unless the transformation from ϕ to $\lambda = (\lambda_1, \dots, \lambda_d)$ is lower triangular, see Datta & Ghosh (1996). Thus the reference posterior has the same invariance property of the fiducial distribution proved in Proposition 1.

Recently Berger et al. (2015) recognize the existence of situations in which one is interested simultaneously in all the parameter components of the model, or in none of them but a prior (and thus a posterior) distribution is necessary to perform other inferences such as predictions. In these cases an *overall* prior is needed. Its determination is an open problem but, as they highlight, when there exists a “common reference prior for all parameters”, this is the natural choice for the overall prior. A similar problem occurs in our context and we will comment on this aspect in the following sections. Notice that here the fiducial distribution (2) suggested by Hannig can be a good choice.

4.1 Location-scale parameter models

For location-scale parameter models the fiducial prior exists and coincides with the reference prior. Assume first that only one parameter, θ , is unknown. In this case the model admits an ancillary statistic \mathbf{Z} and, in particular, we take $Z_i = X_i - X_1$ or $Z_i = X_i/X_1$, $i = 2, \dots, n$, if θ is a location or a scale parameter, respectively.

Proposition 8. *Let $\mathbf{X} = (X_1, \dots, X_n)$ be an i.i.d. sample from a density p_θ , $\theta \in \Theta \subseteq \mathbb{R}$. If θ is a location or a scale parameter, then the fiducial distribution coincides with the Bayesian posterior obtained with the Jeffreys prior $\pi^J(\theta) \propto 1$ or $\pi^J(\theta) \propto 1/\theta$, respectively.*

Example 5. Let \mathbf{X} be an i.i.d. sample from the uniform distribution on $(0, \theta)$, $\theta > 0$, so that θ is a scale parameter. First notice that $S = X_{(n)}$ is a sufficient statistic for θ and thus we can obtain directly the fiducial distribution

$$h_s(\theta) = \frac{\partial}{\partial \theta} H_s(\theta) = -\frac{\partial}{\partial \theta} F_\theta(s) = -\frac{\partial}{\partial \theta} \left(\frac{s}{\theta} \right)^n = \frac{ns^n}{\theta^{n+1}}, \quad \theta > s. \quad (26)$$

However the same result can be obtained without resorting to the sufficient statistic. Set $w = \max(z_2, \dots, z_n)$ and consider the distribution function of X_1 given the ancillary

statistic $\mathbf{Z} = (X_2/X_1, \dots, X_n/X_1)$

$$F_\theta(x_1|\mathbf{z}) = \begin{cases} \left(\frac{x_1}{\theta}\right)^n & 0 < x_1 < \theta, \quad 0 < w \leq 1 \\ \left(\frac{x_1 w}{\theta}\right)^n & 0 < x_1 < \theta, \quad w > 1 \end{cases}. \quad (27)$$

Now, because $w \leq 1$ means $x_1 = \max(x_1, \dots, x_n)$, while for $w > 1$ we have $x_1 w = \max(x_2, \dots, x_n)$, expression (27), as a function of θ , is equivalent to $F_\theta(s)$ appearing in (26) and thus provides the same fiducial distribution. It is immediate to verify that it coincides with the Jeffreys posterior.

A case in which the sufficient statistic is not one-dimensional and thus it is necessary to use an ancillary statistic can be found in the previous Example 3. Trivially $h_s(\theta)$ given in (12) coincides with the Bayesian posterior obtained by $\pi^J(\theta) \propto 1$.

Consider now a model with a location parameter θ and a scale parameter σ , both unknown. Given an i.i.d. sample of size n , an ancillary statistic is, for example, $\mathbf{Z} = (Z_3, \dots, Z_n)$, with $Z_j = (X_j - X_1)/Z_2$, $j = 3, \dots, n$, where $Z_2 = X_2 - X_1$ is marginally ancillary for θ . Then, the one-to-one transformation from \mathbf{X} to (X_1, Z_2, \mathbf{Z}) allows to write the sampling distribution as $p_\sigma(z_2|\mathbf{z})p_\theta(x_1|z_2, \mathbf{z}; \sigma)p(\mathbf{z})$. Note that in specific contexts other transformations could be more appropriate. For example, in a normal model one could use $(\bar{X} = \sum_{i=1}^n X_i/n, S^2 = \sum_{i=1}^n (X_i - \bar{X})^2, \mathbf{Z})$ with $Z_j = (X_j - \bar{X})/S$, $j = 3, \dots, n$, so that the factorization becomes $p_\sigma(s^2|\mathbf{z})p_\theta(\bar{x}|s^2, \mathbf{z}; \sigma)p(\mathbf{z})$.

Proposition 9. *Let $\mathbf{X} = (X_1, \dots, X_n)$ be an i.i.d. sample from a density $p_{\theta, \sigma}$, where θ and σ are a location and a scale parameter, respectively. Then the fiducial distribution $h_{\mathbf{x}}(\sigma, \theta)$ for (σ, θ) coincides with the Bayesian posterior obtained with the reference prior $\pi_{\sigma, \theta}^R(\sigma, \theta) \propto 1/\sigma$.*

Notice that $\pi^R(\sigma, \theta) \propto 1/\sigma$ is different from $\pi^J(\sigma, \theta) \propto 1/\sigma^2$ obtained by the Jeffreys rule which, as already recalled, is not suitable for multidimensional parameters. Furthermore, while π^R does not depend on the ordering of θ and σ , the step-by-step fiducial distribution is in general not allowable if the ordering is reversed. However, $h_{\mathbf{x}}(\sigma, \theta)$ coincides with the fiducial distribution obtained through other “symmetric” approaches, see Hannig (2009) and Fraser (1961). Thus the inferential ordering of importance seems irrelevant for this model and $h_{\mathbf{x}}(\sigma, \theta)$ can be assumed as an overall fiducial distribution.

4.2 Exponential families

Lindley (1958) was the first to study the existence of a fiducial prior, analyzing in particular the case of continuous real NEFs and proving that it exists only for gaussian (with known variance) and gamma (with known shape) models. A full characterization of the real NEFs which admit a fiducial prior is given in V&M (2015). The following proposition summarizes their results.

Proposition 10. *Let \mathcal{F} be a real NEF with natural parameter θ .*

- i) A fiducial prior exists if and only if \mathcal{F} is an affine transformation of one of the following families: normal with known variance, gamma with known shape parameter, binomial, Poisson and negative-binomial. For the three discrete families, the fiducial prior exists for all H_s , H_s^ℓ and H_s^G .*
- ii) When a fiducial prior exists, it belongs to the family of conjugate distributions. Moreover, it coincides with the Jeffreys prior for continuous NEFs and for discrete NEFs too if we choose H_s^G as the fiducial distribution.*
- iii) The fiducial distribution H_s (or H_s^A in the discrete case) and the Bayesian posterior distribution corresponding to the Jeffreys prior have the same Edgeworth's expansion up to the term of order n^{-1} .*

The previous results establish a strong connections between Jeffreys posteriors and fiducial distributions for real NEFs, and thus the two different approaches lead, in some sense, to the same *objective inference*. A discussion about the coverage of the fiducial and the Jeffreys intervals and their good frequentist properties, in particular when compared with the standard Wald intervals, is given in V&M (2015, Section 5).

Consider now a cr-NEF. It is easy to verify that the fiducial distribution $h_{\mathbf{s}}(\phi)$ in (18) belongs to the enriched conjugate family defined in Consonni & Veronese (2001, Section 4.3). This fact is the key-point to prove the following proposition.

Proposition 11. *Let \mathbf{S} be a sufficient statistic distributed according to a cr-NEF on \mathbb{R}^d , parameterized by $\phi = (\phi_1, \dots, \phi_d)$. Then a fiducial prior for ϕ exists if and only if the conditional distribution of S_k given $\mathbf{S}_{[k-1]} = \mathbf{s}_{[k-1]}$ is an affine transformation of one of the following families: normal with known variance, gamma with known shape parameter, binomial, Poisson and negative-binomial.*

In particular, all basic NEF-SQVFs, with the exclusion of the Negative-multinomial/hyperbolic secant, admit a fiducial prior, which belongs to the enriched conjugate family. Moreover, if for the discrete components of these models we consider the geometric mean $h_{\mathbf{s}_{[k]}}^G$, then the product of the Jeffreys priors computed from the conditional distribution of S_k given $\mathbf{S}_{[k-1]} = \mathbf{s}_{[k-1]}$, the reference prior and the fiducial prior are all equal.

Example 4 (ctd.). The multinomial distribution is a basic NEF-SQVF and thus from Proposition 11, setting $s_k = n = 0$ in $h_{\mathbf{s}}^G(\phi)$ given in (20), we obtain the fiducial prior

$$\pi(\phi) \propto \prod_{k=1}^d e^{\phi_k/2} / (1 + e^{\phi_k/2}), \quad (28)$$

which coincides with the reference prior and with the product of the Jeffreys priors for ϕ_k , $k = 1, \dots, d$, computed on the distribution of X_k given $\mathbf{X}_{[k-1]} = \mathbf{x}_{[k-1]}$.

Finally, we observe that the fiducial distribution (17) is always an overall fiducial distribution for ϕ . However, the ϕ -parameterization is often not interesting in itself even if in some cases it is strictly related with a more relevant one. For example, following Berger et al. (2015), consider a multinomial model applied to directional data, as it happens for outcomes from an attitude survey. In this case the cells are naturally ordered, so that it is meaningful to reparameterize the model in terms of the conditional probabilities $p_k^* = \exp(\phi_k)/(1 + \exp(\phi_k))$, $k = 1, \dots, d$. Then $\pi(\phi)$ in (28) induces on $\mathbf{p}^* = (p_1^*, \dots, p_d^*)$ an overall fiducial prior which is a product of independent $\text{Be}(1/2, 1/2)$ distributions coinciding with the overall reference prior.

5 Further examples

5.1 Examples concerning normal models

i) *Difference of means.* Consider two independent normal i.i.d. samples, each of size n , with known common variance σ^2 and means μ_1 and μ_2 , respectively. The sufficient statistics are the sample sums S_1 and S_2 , with $S_i \sim N(n\mu_i, n\sigma^2)$, $i = 1, 2$. If the parameter of interest is $\phi_1 = \mu_2 - \mu_1$, we can reparameterize the joint density of (S_1, S_2) in $(\phi_1 = \mu_2 - \mu_1, \phi_2 = \mu_1)$, so that the conditional distribution of S_2 given $S_1 + S_2$, being $N((n\phi_1 + s_1 + s_2)/2, n\sigma^2/2)$, depends only on ϕ_1 . From Table 1, the fiducial distribution of $\phi_1/2 + (s_1 + s_2)/(2n)$ is $N(s_2/n, \sigma^2/(2n))$, and thus ϕ_1 is $N(\bar{x}_2 - \bar{x}_1, 2\sigma^2/n)$, where $\bar{x}_i = s_i/n$. Because $S_1 + S_2$ is $N(\phi_1 + 2\phi_2, 2n\sigma)$, arguing as before, the fiducial distribution of ϕ_2 given ϕ_1 is $N((\bar{x}_1 + \bar{x}_2 - \phi_1)/2, \sigma^2/(2n))$, so that $h_{S_1, S_2}(\phi_1, \phi_2) = h_{S_1, S_2}(\phi_1)h_{S_1 + S_2}(\phi_2|\phi_1)$. Notice that the same joint fiducial distribution is obtained if we consider the ordering (ϕ_2, ϕ_1) or even if we compute the (marginal) fiducial distributions of μ_1 and μ_2 and obtain that of (ϕ_1, ϕ_2) through the change-of-variable rule. Thus the ordering of the parameter is irrelevant and $h_{S_1, S_2}(\phi_1, \phi_2)$ is an overall fiducial distribution. Furthermore, it coincides with the reference posterior obtained with a constant prior and the marginal distribution of ϕ_1 and ϕ_2 are both confidence distributions.

ii) *Many normal means (Neyman Scott Problem).* Consider n samples of size two (X_{i1}, X_{i2}) , with each X_{ij} independently distributed according to a $N(\mu_i, \sigma^2)$, $i = 1, \dots, n$ and let $\bar{X}_i = (X_{i1} + X_{i2})/2$ and $W = \sum_{i=1}^n (X_{i1} - X_{i2})^2$. The aim is to make inference on the common variance σ^2 , with nuisance parameter $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$. This well known example is used to show that the maximum likelihood estimator $\hat{\sigma}^2 = W/(4n)$ of σ^2 is inconsistent, because $W/(4n) \rightarrow \sigma^2/2$, $n \rightarrow \infty$. To obtain the fiducial distribution of σ^2 , first notice that the joint distribution of the sufficient statistics $\bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_n)$

and W can be factorized as $(\prod_{i=1}^n p_{\mu_i, \sigma^2}(\bar{x}_i)) p_{\sigma^2}(w)$, for the independence of $\bar{\mathbf{X}}$ and W , with $W \sim \text{Ga}(n/2, 1/(4\sigma^2))$. Using Table 1 one can easily obtain from $p_{\sigma^2}(w)$ the fiducial distribution for $1/(4\sigma^2)$, and hence that for σ^2 which is $\text{In-Ga}(n/2, w/4)$, while that of each μ_i given σ^2 , derived from $p_{\mu_i, \sigma^2}(\bar{x}_i)$, is $N(\bar{x}_i, \sigma^2/2)$. As a consequence $h_{\bar{\mathbf{X}}, w}(\sigma^2, \boldsymbol{\mu}) = (\prod_{i=1}^n h_{\bar{x}_i}(\mu_i | \sigma^2)) h_w(\sigma^2)$. This distribution coincides with the posterior obtained from the order invariant reference prior $\pi^R(\sigma^2, \mu_1, \dots, \mu_n) \propto 1/\sigma^2$ and does not present the inconsistency of the likelihood estimator, which instead occurs for the posterior distribution obtained from the Jeffreys prior $\pi^J(\sigma^2, \mu_1, \dots, \mu_n) \propto 1/\sigma^{n+2}$.

5.2 Comparison of two Poisson rates

The comparison of Poisson rates μ_1 and μ_2 is a classical problem arising in many contexts, see for example Lehmann & Romano (2005) for a discussion on an unbiased uniformly most powerful test for the ratio $\phi_1 = \mu_2/\mu_1$. Given two i.i.d. samples of size n from two independent Poisson distributions, the sufficient statistics are the sample sums S_1 and S_2 , with $S_i \sim \text{Po}(n\mu_i)$, $i = 1, 2$. Reparameterizing the joint density of (S_1, S_2) in $(\phi_1 = \mu_2/\mu_1, \phi_2 = \mu_1 + \mu_2)$, we have that the conditional distribution of S_2 given $S_1 + S_2$ is $\text{Bi}(s_1 + s_2, \phi_1/(1 + \phi_1))$ and the marginal distribution of $S_1 + S_2$ is $\text{Po}(n\phi_2)$. Thus the sampling distribution is a cr-NEF and we can apply (16). Using Table 1, the fiducial density for $\phi_1/(1 + \phi_1)$, derived from the conditional distribution of S_2 given $S_1 + S_2$ is $\text{Be}(s_2 + 1/2, s_1 + 1/2)$ which implies

$$h_{s_1, s_2}^G(\phi_1) = \frac{1}{\text{B}(s_2 + 1/2, s_1 + 1/2)} \phi_1^{s_2-1/2} (1 + \phi_1)^{-s_1-s_2-1}, \quad \phi_1 > 0. \quad (29)$$

From the marginal distribution of $S_1 + S_2$ and using again Table 1, it follows that $h_{s_1+s_2}^G(\phi_2)$ is $\text{Ga}(s_1+s_2+1/2, n)$ and thus $h_{s_1, s_2}^G(\phi_1, \phi_2) = h_{s_1, s_2}^G(\phi_1) h_{s_1+s_2}^G(\phi_2)$. This joint fiducial distribution is order-invariant, coincides with the reference posterior according to Proposition 11, and is an overall distribution for (ϕ_1, ϕ_2) . Notice that $h_{s_1, s_2}^G(\phi_1)$ is a confidence distribution and that it differs from the fiducial distribution induced on ϕ_1 by the two independent marginal fiducial densities for μ_1 and μ_2 .

5.3 Bivariate binomial

A Bayesian analysis for the bivariate binomial model has been discussed by Crowder & Sweeting (1989) in connection with a microbiological application. Consider m spores, each with a probability p to germinate, and denote by R the random number of germinating spores, so that R is $\text{Bi}(m, p)$. If q is the probability that one of the latter spores bends in a particular direction and S is the random number of them, the probability distribution of S given $R = r$ is $\text{Bi}(r, q)$. The joint distribution of R and S is called *bivariate binomial*. Crowder and Sweeting observe that the Jeffreys prior $\pi^J(p, q) \propto p^{-1}(1-p)^{-1/2} q^{-1/2}(1-q)^{-1/2}$

is not satisfactory for its asymmetry in p and $1 - p$, while Polson & Wasserman (1990) show that this fact does not occur using the order-invariant reference prior $\pi^R(p, q) \propto p^{-1/2}(1 - p)^{-1/2}q^{-1/2}(1 - q)^{-1/2}$ which is the product of the two independent Jeffreys priors.

The joint fiducial density $h_{r,s}^G(q, p)$ can be obtained as the product of $h_{r,s}^G(q)$, derived from the conditional model $\text{Bi}(r, q)$ of S given $R = r$, and $h_r^G(p|q)$, derived from the marginal model $\text{Bi}(m, p)$ of R , which does not depend on q . Thus p and q are independent under $h_{r,s}^G$ so that it is an overall fiducial distribution. Because for the binomial model the fiducial prior is equal to the Jeffreys prior, see Proposition 10, it follows immediately that $h_{r,s}^G(q, p)$ coincides with the reference posterior. All previous conclusions hold even if we consider the alternative parametrization $(\eta = pq, \lambda = p(1 - q)/(1 - pq))$.

5.4 Ratio of parameters of a trinomial distribution

Bernardo & Ramon (1998) perform the Bayesian reference analysis for the ratio of two multinomial parameters presenting some applications. In particular they discuss the case of (X_1, X_2) distributed according to a trinomial distribution with parameters n and $\mathbf{p} = (p_1, p_2)$, and provide the joint reference prior for $(\phi_1 = p_1/p_2, \phi_2 = p_2)$, with ϕ_1 the parameter of interest. Then they derive the marginal reference posterior for ϕ_1 which is

$$\pi^R(\phi_1|x_1, x_2) \propto \phi_1^{x_1-1/2}(1 + \phi_1)^{-x_1-x_2-1}. \quad (30)$$

To find the fiducial distribution of ϕ_1 , we reparameterize the trinomial model in (ϕ_1, ϕ_2) . The conditional distribution of X_1 given $T = X_1 + X_2 = t$ is $\text{Bi}(t; \phi_1/(1 + \phi_1))$, so that, by Table 1, the fiducial density for $\phi_1/(1 + \phi_1)$ is $\text{Be}(x_1 + 1/2, t - x_1 + 1/2)$ and $h_{x_1,t}^G(\phi_1)$ coincides with (30). From the marginal distribution of T , which is $\text{Bi}(n; \phi_2(1 + \phi_1))$, it is possible to derive the fiducial density

$$h_t^G(\phi_2|\phi_1) = \frac{\Gamma(n + 1)}{\Gamma(t + 1/2)\Gamma(n - t + 1/2)}(1 + \phi_1)^{t+1/2}\phi_2^{t-1/2}(1 - (1 + \phi_1)\phi_2)^{n-t-1/2}$$

so that the joint fiducial density is $h_{x_1,x_2}^G(\phi_1, \phi_2) = h_{x_1,t}^G(\phi_1)h_t^G(\phi_2|\phi_1)$ which coincides with the joint reference posterior.

6 Conclusions and final remarks

We have suggested a way to construct a fiducial distribution which depends on the inferential importance ordering of the parameter components. Our proposal appears to be quite simple to apply and, even if it is not so general as the theory suggested by Hannig, has some advantages in connection with the modern confidence distribution theory and it is strictly related to objective Bayesian analysis.

In complex models an exact analysis is generally not possible, but approximate results can be derived working with asymptotic distributions. In V&M (2015), starting from the sufficient statistic, an expansion up to the first order of the fiducial distribution for the mean parameter of a real NEF is provided. This result can be extended to arbitrary regular models starting from the maximum likelihood estimator of the parameter. When the maximum likelihood estimator is not sufficient a better fiducial distribution can be obtained using an ancillary statistic, as suggested in Section 3. To this aim the magic formula p^* given by Barndorff-Nielsen (1983), which provides an approximation of the conditional distribution of the maximum likelihood estimator given an ancillary statistic, can be fruitfully adopted. Furthermore, these asymptotic results appear to be strictly connected with the theory of matching priors, i.e. priors that ensure approximate frequentist validity of posterior credible set. Notice that also these priors crucially depend on the inferential ordering of the parameters, see Tibshirani (1989) and Datta & Mukerjee (2004). However, a normal approximation of the fiducial distribution, when it can be established and is enough for the analysis, can be proved to be order-invariant. These type of results will be discussed in a forthcoming paper.

Acknowledgements

This research was supported by grants from Bocconi University.

Appendix

A1: Useful results on cr-NEFs

Some technical aspects related to cr-NEFs are the following.

1. A NEF is a cr-NEF if and only if the principal $k \times k$ matrix of the variance function does not depend on μ_{k+1}, \dots, μ_d , for $k = 1, \dots, d - 1$.
2. The Fisher information matrix relative to the ϕ -parametrization is diagonal with the kk -th element depending only on $\phi_{[k]}$.
3. The cumulant transform $M_k(\phi_k; \mathbf{x}_{[k-1]})$ of the k -th conditional density is given by

$$M_k(\phi_k; \mathbf{x}_{[k-1]}) = \sum_{j=1}^{k-1} A_{kj}(\phi_k) x_j + B_k(\phi_k), \quad (31)$$

for some functions A_{kj} and B_k .

4. The conditional expectation of X_k given $\mathbf{X}_{[k-1]} = \mathbf{x}_{[k-1]}$ is linear in $\mathbf{x}_{[k-1]}$, because it is the gradient of (31).

5. The parameter μ_k depends on ϕ only through $\phi_{[k]}$, because from (31)

$$\mu_k = \sum_{j=1}^{k-1} \frac{\partial A_{kj}(\phi_k)}{\partial \phi_k} \mu_j + \frac{\partial B_k(\phi_k)}{\partial \phi_k}. \quad (32)$$

6. Using (14) and (31), it can be checked that

$$\theta_k = \phi_k - \sum_{u=k+1}^d A_{uk}(\phi_u), \quad \text{and} \quad M(\theta) = \sum_{k=1}^d B_k(\phi_k(\theta)). \quad (33)$$

As a consequence of the first part of (33), there exists a function g_k such that

$$\phi_k = \theta_k + g_k(\theta_{k+1}, \dots, \theta_d). \quad (34)$$

Of course all the previous formulas hold for $k = 1, \dots, d$, with the understanding that components that lose meaning for a specific k are set to zero.

A NEF has a *simple quadratic variance function* (SQVF) if the ij -th element of its variance-covariance matrix, seen as a function of the mean parameter $\mu = (\mu_1, \dots, \mu_d)$, can be written as $V_{ij}(\mu) = q\mu_i\mu_j + \sum_{k=1}^d \mu_k L_{ij}^{(k)} + C_{ij}$, where q is a real constant and $L^{(k)}$, $k = 1, \dots, d$ and C are constant $d \times d$ symmetric matrices. Any NEF-SQVF can be obtained, via a nonsingular affine transformation, from one of the *basic* families: Poisson/normal ($q = 0$), multinomial ($q = -1/N$, N positive integer), negative-multinomial ($q = 1/R$, R positive integer), negative-multinomial/gamma/normal ($q = 1/R$) and negative-multinomial/hyperbolic-secant ($q = 1/R$), see Casalis (1996) for a detailed description of these distributions. The ij -th element of the variance function $V(\mu)$ of a basic NEF-SQVF is

$$V_{ij}(\mu) = q\mu_i\mu_j + \sum_{k=1}^d z_{ik}\mu_k + C_{ij}, \quad (35)$$

where $z_{ij} = z_{ji}$, $i, j \in \{1, \dots, d\}$, are constants. The values of z_{ii} for the basic NEF-SQVFs, together with other technical details, are given in the proof of Corollary 2.

A2: Proofs

Proof of Proposition 1.

By the standard change-of-variable rule applied to the first integral in (10), it is enough to show that

$$h_t^\phi(\phi(\lambda)) |J_\phi(\lambda)| = h_t^\lambda(\lambda), \quad (36)$$

where $J_\phi(\boldsymbol{\lambda})$ is the Jacobian of the transformation from ϕ to $\boldsymbol{\lambda}$. Now, from (8) we have

$$\begin{aligned} h_{\mathbf{t}}^\phi(\phi(\boldsymbol{\lambda})) &= \prod_{k=1}^d h_{\mathbf{t}_{[k]}, \mathbf{t}_{-[d]}}(\phi_{d-k+1}(\lambda_{[d-k+1]}) | \phi_{[d-k]}(\boldsymbol{\lambda}_{[d-k]})) \\ &= \prod_{k=1}^d \left| \frac{\partial}{\partial \phi_{d-k+1}} F_{\phi_{d-k+1}}(t_k | \mathbf{t}_{[k-1]}, \mathbf{t}_{-[d]}; \phi_{[d-k]}) \right| \Big|_{\phi_{[d-k+1]} = \phi_{[d-k+1]}(\boldsymbol{\lambda}_{[d-k+1]})} \end{aligned}$$

while

$$J_\phi(\boldsymbol{\lambda}) = \prod_{k=1}^d \frac{\partial \phi_{d-k+1}(\boldsymbol{\lambda}_{[d-k+1]})}{\partial \lambda_{d-k+1}},$$

because the transformation from ϕ to $\boldsymbol{\lambda}$ is lower triangular. It follows from the last two formulas and the chain rule that

$$h_{\mathbf{t}}^\phi(\phi(\boldsymbol{\lambda})) | J_\phi(\boldsymbol{\lambda})| = \prod_{k=1}^d \left| \frac{\partial}{\partial \lambda_{d-k+1}} F_{\lambda_{d-k+1}}(t_k | \mathbf{t}_{[k-1]}, \mathbf{t}_{-[d]}; \boldsymbol{\lambda}_{[d-k]}) \right|,$$

where $F_{\lambda_{d-k+1}}$ is the distribution function of T_k given $(\mathbf{T}_{[k-1]} = \mathbf{t}_{[k-1]}, \mathbf{T}_{[-d]} = \mathbf{t}_{[-d]})$ in the $\boldsymbol{\lambda}$ parameterization. The equality (36) follows by applying (9) to the model parameterized by $\boldsymbol{\lambda}$. \diamond

Proof of Proposition 2.

Let $p^G(x) = c^{-1}(p_1(x)p_2(x))^{1/2}$, where $c = \int (p_1(x)p_2(x))^{1/2} d\nu(x)$ is the normalizing constant. Then

$$\begin{aligned} KL(q|p_1) + KL(q|p_2) &= \int \left(\log \frac{q(x)}{p_1(x)} \right) q(x) d\nu(x) + \int \left(\log \frac{q(x)}{p_2(x)} \right) q(x) d\nu(x) \\ &= \int \left(\log \frac{q^2(x)}{p_1(x)p_2(x)} \right) q(x) d\nu(x) = 2 \int \left(\log \frac{q(x)}{C p^G(x)} \right) q(x) d\nu(x) \\ &= 2 \int \left(\log \frac{q(x)}{p^G(x)} \right) q(x) d\nu(x) - 2 \log c = 2KL(q|p^G) - 2 \log c. \end{aligned} \quad (37)$$

Because c does not depend on q , it follows that the functional in (37) achieves its minimum (equal to $-2 \log c$) if and only if $KL(q|p^G) = 0$, i.e. $q = p^G$. \diamond

Proof of Proposition 3.

We only prove that $H_s(\theta) < H_s^G(\theta)$; the other inequality can be shown in the same way. Using (5) and (6), we can write

$$\begin{aligned} \frac{h_s^G(\theta)}{h_s(\theta)} &= \frac{1}{c} \frac{\sqrt{h_s(\theta) h_s^\ell(\theta)}}{h_s(\theta)} = \frac{1}{c} \frac{\sqrt{h_s(\theta) (h_s(\theta) + \frac{\partial}{\partial \theta} p_\theta(s))}}{h_s(\theta)} \\ &= \frac{1}{c} \sqrt{1 + \frac{\frac{\partial}{\partial \theta} p_\theta(s)}{h_s(\theta)}} = \frac{1}{c} \sqrt{1 + \gamma_s(\theta)}. \end{aligned} \quad (38)$$

By hypothesis $\gamma_s(\theta)$ is decreasing and thus, from (38), $h_s^G(\theta)/h_s(\theta)$ is also decreasing on Θ . This is a sufficient condition for $H_s(\theta) < H_s^G(\theta)$, see Shaked & Shanthikumar (2007, Theorem 1.C.1). \diamond

Proof of Corollary 1.

Let $p_\theta(s) = \exp(\theta s - nM(\theta))$ be the probability mass function (with respect to a measure ν) of a real NEF, with θ the natural parameter. Fixing $s \in \mathcal{S}_0$, we can write

$$\begin{aligned} \gamma_s(\theta) &= \left(\frac{\partial p_\theta(s)}{\partial \theta} \right) / \left(\frac{\partial}{\partial \theta} (1 - F_\theta(s)) \right) = \frac{(s - nM'(\theta)) \exp(\theta s - nM(\theta))}{\sum_{t=s+1}^{+\infty} (t - nM'(\theta)) \exp(\theta t - nM(\theta))} \\ &= \left(\sum_{t=s+1}^{+\infty} \frac{t - nM'(\theta)}{s - nM'(\theta)} \exp\{(t - s)\theta\} \right)^{-1}. \end{aligned} \quad (39)$$

The elements in the sum (39) are continuous and increasing functions of θ in both intervals for which $\theta < \widehat{\theta}_s$ and $\theta > \widehat{\theta}_s$, where $\widehat{\theta}_s = (M')^{-1}(s/n)$. Thus $\gamma_s(\theta)$ is decreasing in these intervals. Moreover, $\gamma_s(\theta)$ is equal to zero for $\theta = \widehat{\theta}_s$, positive for $\theta < \widehat{\theta}_s$ and negative for $\theta > \widehat{\theta}_s$, because the denominator of $\gamma_s(\theta)$ is $h_s(\theta)$, which is positive. Then $\gamma_s(\theta)$ is decreasing over all Θ and from Proposition 3 the result follows. \diamond

Proof of Proposition 4.

In order to prove the proposition, it is sufficient to show that there exist θ_1 and θ_2 in Θ , $\theta_1 < \theta_2$, such that $h_s^A(\theta_i) = h_s^G(\theta_i)$, $i = 1, 2$, with $h_s^A(\theta) < h_s^G(\theta)$ for $\theta_1 < \theta < \theta_2$ and $h_s^A(\theta) > h_s^G(\theta)$ otherwise, see Shaked & Shanthikumar (2007, proof of Theorem 3.A.44). Thus we analyze the sign of $h_s^A(\theta) - h_s^G(\theta)$. We can write

$$\begin{aligned} h_s^A(\theta) - h_s^G(\theta) &= \frac{1}{2}(h_s(\theta) + h_s^\ell(\theta)) - \frac{1}{c} \sqrt{h_s(\theta) h_s^\ell(\theta)} \\ &= h_s(\theta) \left(\frac{1}{2}(2 + \gamma_s(\theta)) - \frac{1}{c} \sqrt{1 + \gamma_s(\theta)} \right), \end{aligned}$$

so that the sign of the difference $h_s^A(\theta) - h_s^G(\theta)$ is a function of $\gamma_s(\theta)$ only. First notice that, by a standard property of the arithmetic and geometric means, $c = \int \sqrt{h_s(\theta) h_s^\ell(\theta)} d\theta < \int (h_s(\theta) + h_s^\ell(\theta))/2 d\theta = 1$ for all θ . After some straightforward algebra, it can be seen that $h_s^A(\theta) - h_s^G(\theta) = 0$ when (and only when) $\gamma_s(\theta) = 2c^{-2}((1 - c^2) - \sqrt{1 - c^2}) = k_1$ or $\gamma_s(\theta) = 2c^{-2}((1 - c^2) + \sqrt{1 - c^2}) = k_2$, with $k_1 \in (-1, 0)$ and $k_2 > 0$. Moreover we have $h_s^A(\theta) < h_s^G(\theta)$ for $k_1 < \gamma_s(\theta) < k_2$ and $h_s^A(\theta) > h_s^G(\theta)$ for $\gamma_s(\theta) < k_1$ or $\gamma_s(\theta) > k_2$. By assumption, $\gamma_s(\theta)$ is decreasing on Θ from $+\infty$ to -1 , so that there exist θ_1 and θ_2 , with $\gamma_s(\theta_1) = k_2$ and $\gamma_s(\theta_2) = k_1$, satisfying the sufficient condition stated at the beginning of the proof. \diamond

Proof of Proposition 5.

First notice that if we use for constructing the fiducial distribution the sufficient statistic S , which has the same dimension of the parameter, we do not need an ancillary statistic. Furthermore, \mathbf{T} is a one-to-one transformation of \mathbf{X} , and thus \mathbf{S} is a function of $\mathbf{T} = (\mathbf{T}_{[d]}, \mathbf{T}_{-[d]})$ but, since \mathbf{S} is complete, it is stochastically independent of $\mathbf{T}_{-[d]}$ by Basu's Theorem. As a consequence, $\mathbf{S}_{[k]}$ is also independent of $\mathbf{T}_{-[d]}$ and thus

$$\begin{aligned} & \Pr_{\phi_{d-k+1}}(S_k \leq s_k \mid \mathbf{S}_{[k-1]} = \mathbf{s}_{[k-1]}; \phi_{[d-k]}) = \\ & \Pr_{\phi_{d-k+1}}(S_k \leq s_k \mid \mathbf{S}_{[k-1]} = \mathbf{s}_{[k-1]}, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]}; \phi_{[d-k]}). \end{aligned} \quad (40)$$

From the one-to-one lower triangular transformation $\mathbf{s} = \mathbf{g}(\mathbf{t}_{[d]}, \mathbf{t}_{-[d]})$, we have that $s_k = g_k(t_k, \mathbf{t}_{[k-1]}, \mathbf{t}_{-[d]})$, with g_k invertible with respect to t_k , so that (assuming g_k increasing) (40) becomes

$$\begin{aligned} & \Pr_{\phi_{d-k+1}}(g_k(T_k, \mathbf{T}_{[k-1]}, \mathbf{T}_{-[d]}) \leq s_k \mid \mathbf{T}_{[k-1]} = \mathbf{t}_{[k-1]}, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]}; \phi_{[d-k]}) = \\ & \Pr_{\phi_{d-k+1}}(T_k \leq g_k^{-1}(s_k, \mathbf{T}_{[k-1]}, \mathbf{T}_{-[d]}) \mid \mathbf{T}_{[k-1]} = \mathbf{t}_{[k-1]}, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]}; \phi_{[d-k]}) = \\ & \Pr_{\phi_{d-k+1}}(T_k \leq t_k \mid \mathbf{T}_{[k-1]} = \mathbf{t}_{[k-1]}, \mathbf{T}_{-[d]} = \mathbf{t}_{-[d]}; \phi_{[d-k]}), \end{aligned}$$

which proves the proposition. \diamond

Proof of Proposition 6.

Because each conditional distribution of X_k given $\mathbf{X}_{[k-1]} = \mathbf{x}_{[k-1]}$ belongs to a NEF with natural parameter ϕ_k , using (4) we have that $H_{\mathbf{s}_{[k]}}(\phi_k)$ is a distribution function for ϕ_k . The result follows from the postulated independence among the ϕ_k 's. \diamond

Proof of Proposition 7.

Formulas (18) and (19) derive by a direct application of (17) to the conditional distributions of the different families. For a detailed description of the cr-NEFs involved, see Consonni & Veronese (2001, proof of Theorem 3). \diamond

Proof of Corollary 2.

First notice that the fiducial distribution $h_{\mathbf{s}}(\boldsymbol{\mu})$ can be more easily obtained via a double transformation, namely

$$h_{\mathbf{s}}^{\boldsymbol{\mu}}(\boldsymbol{\mu}) = h_{\mathbf{s}}^{\boldsymbol{\theta}}(\boldsymbol{\theta}(\boldsymbol{\mu})) |J_{\boldsymbol{\phi}}(\boldsymbol{\theta}(\boldsymbol{\mu}))| |J_{\boldsymbol{\theta}}(\boldsymbol{\mu})|,$$

where the Jacobian $|J_{\boldsymbol{\phi}}(\boldsymbol{\theta})| = 1$ for (34), and

$$|J_{\boldsymbol{\theta}}(\boldsymbol{\mu})| \propto \det\{\mathbf{V}(\boldsymbol{\mu})\}^{-1} = \exp \left\{ - \sum_{k=1}^d \theta_k(\boldsymbol{\mu}) z_k - q(d+1)M(\boldsymbol{\theta}(\boldsymbol{\mu})) \right\}, \quad (41)$$

see Gutiérrez-Peña & Smith (1997, pag. 34) for the proportionality relationship and Consonni et al. (2004, Prop. 1) for the equality. We consider now each family.

Poisson/normal family with m Poisson components. We have $q = 0$, $\phi_k = \theta_k$; $z_k = 1$, $\theta_k = \log(\mu_k)$ and $B_k(\phi_k) = \exp(\phi_k)$ for $k = 1, \dots, m$, while $z_k = 0$, $\theta_k = \mu_k/\sigma^2$ and $B_k(\phi_k) = \sigma^2 \phi_k^2/2$ for $k = m+1, \dots, d$, where σ^2 is the known variance of the normal components. Then from (41), it follows that $|J_\phi(\boldsymbol{\theta}(\boldsymbol{\mu}))| = \prod_{k=1}^m \mu_k^{-1}$, and thus using (18), the result (21) follows.

Multinomial family. Using the relationships in Example 1, (41) gives $|J_\phi(\boldsymbol{\theta}(\boldsymbol{\mu}))| = (N - \sum_{k=1}^d \mu_k)^{-1} \prod_{k=1}^d \mu_k^{-1}$ and using (8) we obtain (22).

Negative-multinomial family. We have $q = 1/R$, $R > 0$, $z_k = 1$, $\phi_k(\boldsymbol{\theta}) = \theta_k - \log(1 - \sum_{u=k+1}^d e^{\theta_u})$, $\theta_k = \log(\mu_k) - \log(R + \sum_{j=1}^d \mu_j)$, and $B_k(\phi_k) = -R \log(1 - e^{\phi_k})$ for $k = 1, \dots, d$. Then from (41), it follows that $|J_\phi(\boldsymbol{\theta}(\boldsymbol{\mu}))| = (R + \sum_{k=1}^d \mu_k)^{-1} \prod_{k=1}^d \mu_k^{-1}$, and thus using (18), the result (23) follows.

Negative-multinomial/gamma/normal family (with an m dimensional negative-multinomial component). We have $q = 1/R$, $R > 0$; $z_k = 1$ and $\phi_k = \log(\mu_k/(R + \sum_{j=1}^k \mu_j))$, $k = 1, \dots, m$; $z_{m+1} = 0$ and $\phi_{m+1} = -(R + \sum_{j=1}^m \mu_j)/\mu_{m+1}$; $z_k = 0$ and $\phi_k = \mu_k/\mu_{m+1}$, $k = m+2, \dots, d$. In this case it is convenient to compute the fiducial density of $\boldsymbol{\mu}$ directly from (19). Observing that the Jacobian of the transformation from $\boldsymbol{\phi}$ to $\boldsymbol{\mu}$ is

$$|J_\phi(\boldsymbol{\mu})| = \left(R + \sum_{k=1}^m \mu_k\right)^{-1} \left(\prod_{k=1}^m \mu_k^{-1}\right) \left(R + \sum_{j=1}^m \mu_j\right) \mu_{m+1}^{-2} \mu_{m+1}^{-d+m+1}$$

and using the previous expression of ϕ_k , the density (24) follows. \diamond

Proof of Proposition 8.

Let \mathbf{X} be an i.i.d. sample of size n , with $X_i \sim p_\theta(x_i) = f(x_i - \theta)$, i.e. θ is a location parameter, and consider the transformation $Z_1 = X_1$, $Z_i = X_i - X_1$, $i = 2, \dots, n$, whose Jacobian is one. Then, setting $\mathbf{z} = (z_1, \dots, z_n)$

$$H_{\mathbf{x}}(\theta) = H_{\mathbf{z}}(\theta) = 1 - F_\theta(z_1|z_2, \dots, z_n) = \frac{\int_{z_1}^{+\infty} f(t - \theta) \prod_{i=2}^n f(t + z_i - \theta) dt}{\int_{-\infty}^{+\infty} f(t - \theta) \prod_{i=2}^n f(t + z_i - \theta) dt}.$$

Using now the substitution $m = -t + \theta + z_1$ in the previous two integrals, and recalling that $\pi^J(\theta) \propto 1$ we obtain

$$\begin{aligned} \frac{\int_{-\infty}^{\theta} f(z_1 - m) \prod_{i=2}^n f(z_1 + z_i - m) dm}{\int_{-\infty}^{+\infty} f(z_1 - m) \prod_{i=2}^n f(z_1 + z_i - m) dm} &= \frac{\int_{-\infty}^{\theta} \prod_{i=1}^n f(x_i - m) \pi^J(m) dm}{\int_{-\infty}^{+\infty} \prod_{i=1}^n f(x_i - w) \pi^J(m) dm} \\ &= \int_{-\infty}^{\theta} \pi^J(m|\mathbf{x}) dm. \end{aligned}$$

The result relative to the scale parameter follows recalling that the model $p_\theta(x) = f(x/\theta)/\theta$ can be transformed in a model with location parameter μ setting $y = \log(x)$ and $\mu = \log(\theta)$. In this case a constant prior on μ is equivalent to a prior on θ proportional to $1/\theta$. \diamond

Proof of Proposition 9.

Let \mathbf{X} be an i.i.d. sample of size n , with $X_i \sim p_{\theta,\sigma}(x_i) = f((x_i - \theta)/\sigma)/\sigma$, $i = 1, \dots, n$ and notice that the absolute value of the Jacobian of the transformation from \mathbf{x} to (x_1, z_2, \mathbf{z}) , with $z_2 = x_2 - x_1$, $\mathbf{z} = (z_3, \dots, z_n)$ and $z_j = (x_j - x_1)/z_2$, $j = 3, \dots, n$, is $|z_2|^{n-2}$. Furthermore, the reference prior $\pi^R(\theta, \sigma)$ is order-invariant and can be written as $\pi^R(\theta|\sigma)\pi^R(\sigma)$, where $\pi^R(\theta|\sigma) \propto 1$ and $\pi(\sigma) \propto 1/\sigma$, see Fernández & Steel (1999). Working conditionally on σ we can thus apply Proposition 8 to conclude that the reference posterior and the fiducial distribution for θ given σ coincide. It remains to show that

$$\begin{aligned} \pi^R(\sigma|\mathbf{x}) &= \pi^R(\sigma|x_1, z_2, \mathbf{z}) \propto \int_{-\infty}^{\infty} p_{\theta,\sigma}(x_1, z_2, \mathbf{z}) \pi^R(\theta|\sigma) \pi^R(\sigma) d\theta \\ &\propto \int_{-\infty}^{\infty} \frac{1}{\sigma^{n+1}} f\left(\frac{x_1 - \theta}{\sigma}\right) f\left(\frac{z_2 + x_1 - \theta}{\sigma}\right) \prod_{i=3}^n f\left(\frac{z_2 z_i + x_1 - \theta}{\sigma}\right) d\theta \end{aligned} \quad (42)$$

corresponds to the fiducial density $h_{z_2, \mathbf{z}}(\sigma)$.

We have

$$\begin{aligned} H_{z_2, \mathbf{z}}(\sigma) &= 1 - F_\sigma(z_2|\mathbf{z}) = \int_{z_2}^{+\infty} \int_{-\infty}^{+\infty} p_{\theta,\sigma}^{X_1, Z_2|\mathbf{Z}}(t, w|\mathbf{z}) / p^{\mathbf{Z}}(\mathbf{z}) dt dw \\ &= \frac{\int_{z_2}^{+\infty} \int_{-\infty}^{+\infty} \frac{|w|^{n-2}}{\sigma^n} f\left(\frac{t-\theta}{\sigma}\right) f\left(\frac{w+t-\theta}{\sigma}\right) \prod_{i=3}^n f\left(\frac{wz_i+t-\theta}{\sigma}\right) dt dw}{p^{\mathbf{Z}}(\mathbf{z})}, \end{aligned} \quad (43)$$

where the density of $p^{\mathbf{Z}}(\mathbf{z})$ does not depend on the parameters because \mathbf{z} is ancillary. Assuming $z_2 > 0$, and using the transformation $m = x_1 - v(t - \theta)/\sigma$, $v = z_2\sigma/w$, which implies $t = \sigma(x_1 - m)/v + \theta$, $w = z_2\sigma/v$, with Jacobian $z_2\sigma^2/v^3$, the fiducial distribution $H_{z_2, \mathbf{z}}(\sigma)$ in (43) becomes

$$\frac{\int_0^\sigma \int_{-\infty}^{+\infty} \frac{z_2^{n-1}}{v^{n+1}} f\left(\frac{x_1-m}{v}\right) f\left(\frac{z_2+x_1-m}{v}\right) \prod_{i=3}^n f\left(\frac{z_2 z_i + x_1 - m}{v}\right) dm dv}{p^{\mathbf{Z}}(\mathbf{z})}.$$

Taking the derivative with respect to σ , it is immediate to see that the fiducial density for σ coincides with the posterior distribution given in (42).

If $z_2 < 0$, and applying to the integral the same transformation used in the previous case, we have

$$\begin{aligned} F_\sigma(z_2|\mathbf{z}) &= \frac{\int_{-\infty}^{z_2} \int_{-\infty}^{+\infty} \frac{|w|^{n-2}}{\sigma^n} f\left(\frac{t-\theta}{\sigma}\right) f\left(\frac{w+t-\theta}{\sigma}\right) \prod_{i=3}^n f\left(\frac{wz_i+t-\theta}{\sigma}\right) dt dw}{p^{\mathbf{Z}}(\mathbf{z})} \\ &= - \frac{\int_0^\sigma \int_{-\infty}^{+\infty} \frac{(-z_2)^{n-1}}{v^{n+1}} f\left(\frac{x_1-m}{v}\right) f\left(\frac{z_2+x_1-m}{v}\right) \prod_{i=3}^n f\left(\frac{z_2 z_i + x_1 - m}{v}\right) dm dv}{p^{\mathbf{Z}}(\mathbf{z})}, \end{aligned}$$

so that again the derivative with respect to σ of $H_{z_2, \mathbf{z}}(\sigma) = 1 - F_\sigma(z_2 | \mathbf{z})$ leads to (42). \diamond

The following lemma will be used in the proof of Proposition 11.

Lemma 1. *Consider a cr-NEF on \mathbb{R}^d , with the k -th diagonal element in the Fisher information matrix given by $I_{kk}(\phi) = a_k(\phi_k)b_k(\phi_{[k-1]})$. Then the d -group (order-invariant) reference prior π^R for $\phi = (\phi_1, \dots, \phi_d)$ is*

$$\pi^R(\phi) = \prod_{k=1}^d \pi_k^J(\phi_k) \propto \prod_{k=1}^d (a_k(\phi_k))^{1/2}, \quad (44)$$

where $\pi_k^J(\phi_k)$ is the Jeffreys prior obtained from the conditional distribution of X_k given $\mathbf{X}_{[k-1]} = \mathbf{x}_{[k-1]}$.

Proof of Lemma 1

First observe that $\boldsymbol{\mu}_{[k]}$ is a one-to-one transformation of $\phi_{[k]}$ and that the information matrix $\mathbf{I}(\phi)$ of a cr-NEF is diagonal, see Appendix A1 (points 2 and 5). From (14) and (31), the kk -th element of \mathbf{I} is

$$\begin{aligned} I_{kk}(\phi_{[k]}) &= -E_{\phi}^{\mathbf{X}} \left(\frac{\partial^2}{\partial \phi_k^2} \log p_{\phi_k}(x_k | \mathbf{x}_{[k-1]}; \phi_k) \right) = -E_{\phi}^{\mathbf{X}} (M_k''(\phi_k; \mathbf{x}_{[k-1]})) \\ &= \sum_{j=1}^{k-1} A_{kj}''(\phi_k) \mu_j(\phi_{[j]}) + B_k''(\phi_k). \end{aligned}$$

Under the assumption in the proposition, we can write $I_{kk}(\phi_{[k]}) = a_k(\phi_k)b_k^*(\boldsymbol{\mu}_{[k-1]}(\phi_{[k-1]}))$. From Datta & Ghosh (1995), it follows that the reference prior on ϕ is order-invariant and is given by the last product in (44).

Consider now the Jeffreys prior on ϕ_k obtained from $p_{\phi_k}(x_k | \mathbf{x}_{[k-1]})$. This is proportional to the square root of

$$-E^{X_k | \mathbf{x}_{[k-1]}} (M_k''(\phi_k; \mathbf{x}_{[k-1]})) = \sum_{j=1}^{k-1} A_{kj}''(\phi_k) x_j + B_k''(\phi_k) = a_k(\phi_k)b_k^*(\mathbf{x}_{[k-1]}),$$

where again the last equality holds by the assumption in the proposition. Thus the product of the d Jeffreys priors is equal to (44) and the result holds. \diamond

Proof of Proposition 11.

Due to the independence of the ϕ_k 's, a fiducial prior for ϕ exists if and only if there exists a fiducial prior for each ϕ_k . Because the conditional distribution of S_k given $\mathbf{S}_{[k-1]} = \mathbf{s}_{[k-1]}$ belongs to a real NEF with natural parameter ϕ_k , the result of the first part of the proposition follows from Proposition 10.

The first statement of the second part of the proposition follows checking directly the form of the conditional distributions of the basic NEF-SQVFs and using again Proposition 10. The second statement follows from the remark stated before the proposition and from Lemma 1. \diamond

References

- Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365.
- Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis* **1**, 385–402.
- Berger, J. O. & Bernardo, J. M. (1992). Ordered group reference priors with application to a multinomial problem. *Biometrika* **79**, 25–37.
- Berger, J. O. Bernardo, J. M. & Sun, D. (2015). Overall objective priors. *Bayesian Analysis* **10**, 189–221.
- Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. R. Stat. Soc. Ser. B* **41**, 113–147.
- Bernardo, J.M. & Ramon, J.M. (1998). An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *The Statistician* **28**, 101–135.
- Bernardo, J. M. & Smith, A. F. M. (1994). *Bayesian Theory*. Wiley: Chichester.
- Casalis, M. (1996). The $2d+4$ simple quadratic natural exponential families on \mathbb{R}^d . *Ann. Statist.* **24**, 1828–1854.
- Consonni, G. & Veronese, P. (2001). Conditionally reducible natural exponential families and enriched conjugate priors. *Scand. J. Stat.* **28**, 377–406.
- Consonni, G., Veronese, P. & Gutiérrez-Peña, E. (2004). Reference priors for exponential families with simple quadratic variance function. *J. Multivariate Anal.* **88**, 335–364.
- Crowder, M., & Sweeting, T. (1989). Bayesian inference for a bivariate binomial distribution. *Biometrika* **76**, 599–603.
- Datta, G. S. & Ghosh, M. (1995). Some Remarks on Noninformative Priors. *J. Amer. Statist. Assoc.* **90**, 1357–1363.
- Datta, G. S. & Ghosh, M. (1996). On the Invariance of Noninformative Priors. *Ann. Statist.* **24**, 141–159.
- Datta, G. S. & Mukerjee, R. (2004). *Probability matching priors: higher order asymptotics (Lecture Notes in Statistics)*. Springer: New York.
- Dawid, A. P. & Stone, M. (1982). The functional-model basis of fiducial inference. *Ann. Statist.* **10**, 1054–1074.

- Dempster, A. P. (1963). Further examples of inconsistencies in the fiducial argument. *Ann. Statist.* **34**, 884–891.
- Fernández, C. & Steel, F. J. (1999). Reference priors for the general location-scale model. *Statist. Prob. Lett.* **43**, 377–384.
- Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society* **26**, 4, 528–535.
- Fisher, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugenics* **VI**, 91–98.
- Fisher, R. A. (1973). *Statistical methods and scientific inference*. Hafner Press: New York.
- Fraser, D. A. S. (1961). On fiducial inference. *Ann. Math. Statist.* **32**, 661–676.
- Gutiérrez-Peña, E. & Smith, A. F. M. (1997). Exponential and Bayesian conjugate families: review and extensions (with discussion). *Test* **6**, 1–90.
- Hannig, J. (2009). On generalized fiducial inference. *Statist. Sinica* **19**, 491–544.
- Hannig, J. (2013). Generalized fiducial inference via discretization. *Statist. Sinica* **23**, 489–514.
- Hannig, J. & Iyer, H. (2008). Fiducial intervals for variance components in an unbalanced two-component normal mixed linear model. *J. Amer. Statist. Assoc.* **103**, 854–865.
- Hannig, J., Iyer, H. K. & Wang, C. M. (2007). Fiducial approach to uncertainty assessment accounting for error due to instrument resolution. *Metrologia* **44**, 476–483.
- Hannig, J., Iyer, H. K., Lai, R. C. S. & Lee T. C. M. (2016). Generalized Fiducial Inference: A Review and New Results. *J. American Statist. Assoc.* **44**, 476–483.
- Johnson, N. L., Kemp, W. A. & Kotz, J. P. (2005). *Univariate discrete distributions*. Wiley: New York.
- Krishnamoorthy, K. & Lee, M. (2010). Inference for functions of parameters in discrete distributions based on fiducial approach: binomial and Poisson cases. *J. Statist. Plann. Inference.* **140**, 1182–1192.
- Lehmann, E. L. & Romano, J. P. (2005). *Testing statistical hypotheses*. Springer: New York.

- Lindley, D. V. (1958). Fiducial distributions and Bayes theorem. *J. R. Stat. Soc. Ser. B* **20**, 102–107.
- Martin, R. & Liu, C. (2013). Inferential models: a framework for prior-free posterior probabilistic inference. *J. Amer. Statist. Assoc.* **108**, 301–313.
- Petrone, S. & Veronese, P. (2010). Feller operators and mixture priors in Bayesian nonparametrics. *Statist. Sinica* **20**, 379–404.
- Polson, N., & Wasserman, L. (1990). Prior distributions for the bivariate binomial. *Biometrika* **77**, 901–904.
- Schweder, T. & Hjort, N. L. (2002). Confidence and likelihood. *Scand. J. Stat.* **29**, 309–332.
- Schweder, T. & Hjort, N. L. (2016). *Confidence, likelihood and probability*. London: Cambridge University Press.
- Shaked, M. & Shanthikumar, J. G. (2007). *Stochastic orders*. Springer: New York.
- Singh, K., Xie, M. & Strawderman, M. (2005). Combining information through confidence distribution. *Ann. Statist.* **33**, 159–183.
- Stein, C. (1959). *An example of wide discrepancy between fiducial and confidence intervals*. *Ann. Math. Statist.* **30**, 877–880.
- Taraldsen, G. & Lindqvist, B. H. (2013). Fiducial theory and optimal inference. *Ann. Statist.* **41**, 323–341.
- Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604–608.
- Veronese, P. & Melilli, E. (2015). Fiducial and confidence distributions for real exponential families. *Scand. J. Stat.* **42**, 471–484.
- Wandler, D. & Hannig, J. (2012). A fiducial approach to multiple comparisons. *J. Statist. Plann. Inference* **142**, 878–895.
- Wilkinson, G. N. (1977). On resolving the controversy in statistical inference. *J. R. Stat. Soc. Ser. B* **39**, 119–171.